Politecnico
di Torino

Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Matematica

A.a. 2024/2025

Sessione di Laurea ottobre 2025

Stima dell'efficacia vaccinale

Un approccio bayesiano generale per un'inferenza più precisa in studi clinici con campioni limitati

Relatori:

Prof. Mauro Gasparini

Dr. Marco Ratta

Candidato:

Vincenzo Di Trani

Ringraziamenti

Desidero esprimere la mia più sincera gratitudine al Dr. Ratta per la disponibilità, la competenza e la preziosa guida durante la stesura di questa tesi.
Ringrazio inoltre il Prof. Gasparini per i suggerimenti e l'attenzione riservata al mio lavoro.
Infine, un grazie speciale va alla mia famiglia per la pazienza e il supporto durante questo percorso.

Sommario

In questo lavoro si illustra un metodo volto a migliorare la stima dell'efficacia vaccinale (VE), una misura ampiamente utilizzata nella ricerca clinica sui vaccini, in presenza di campioni di dimensioni piccole o medie. Si introduce un approccio bayesiano esaustivo che migliora le metodologie esistenti, considerando esplicitamente i processi di reclutamento dei pazienti per informare la stima parametrica. In particolare, a differenza della maggior parte dei metodi attualmente utilizzati, nella metodologia proposta sia il numero di infezioni che i tempi di sorveglianza censurati sono trattati come statistiche informative, le cui distribuzioni sottostanti sono impiegate per derivare la verosimiglianza completa. Il modello risultante dipende dai primi due momenti dei tempi di sorveglianza e, per dimensioni campionarie finite, migliora sia il metodo frequentista di massima verosimiglianza, che si basa esclusivamente sul primo momento, sia il metodo bayesiano beta-binomiale utilizzato nello studio clinico sul vaccino anti-Covid-19 sponsorizzato da Pfizer/BioNTech nel 2020. La metodologia proposta è validata tramite ampie simulazioni numeriche, che dimostrano sostanziali miglioramenti nella precisione della stima del parametro di interesse in scenari diversi e sotto molteplici piani di reclutamento, quando il numero di eventi (e approssimativamente la numerosità campionaria) è piccolo o medio. Per grandi campioni, questo approccio risulta equivalente alla massima verosimiglianza. Le simulazioni Monte Carlo tramite catene di Markov (MCMC) sono necessarie, ma possono essere condotte in maniera molto efficiente grazie a una parametrizzazione appropriata.

Indice

Parte I

Capitolo 1

Studi clinici

1.1 Introduzione agli studi clinici

In questo primo capitolo vengono introdotti gli studi clinici, con l'obiettivo di fornire al lettore il contesto generale e i concetti di base che saranno ripresi e applicati nella seconda parte della tesi. Per un approfondimento più esteso e dettagliato sull'argomento, si può fare riferimento a [25].
Uno *studio clinico controllato* (*trial clinico*) è uno studio prospettico e comparativo che valuta gli effetti e il valore terapeutico (o l'utilità clinica) di uno o più interventi sperimentali somministrati a un gruppo di partecipanti, denominato *gruppo di intervento*, confrontandoli con quelli osservati in un gruppo di pazienti che non ricevono alcun trattamento (placebo) oppure ai quali viene somministrato un trattamento standard, denominato *gruppo di controllo*. I gruppi di pazienti che partecipano allo studio sono anche definiti *bracci* dello studio.
Si noti che uno studio clinico è prospettico e non retrospettivo. In uno studio prospettico i partecipanti vengono seguiti nel tempo a partire da un momento iniziale ben definito, cioè prima che si verifichino gli eventi di interesse (insorgenza di una malattia ad esempio). L'obiettivo è osservare cosa succede dopo, registrando se e quando si sviluppano certe condizioni o effetti in seguito a un trattamento. In questo senso, lo studio prospettico avviene in "avanti nel tempo": si parte dal presente e si va verso il futuro, registrando i dati in tempo reale. In uno studio retrospettivo, invece, si parte dalla fine: si individuano soggetti che hanno già sperimentato o meno l'evento e si va indietro nel tempo per ricostruirne le possibili cause. I dati vengono presi da cartelle cliniche,

registri sanitari o interviste, senza un'osservazione diretta nel tempo. Dunque, lo studio retrospettivo è "all'indietro nel tempo": si parte dal presente e si analizza il passato.
Le tecniche di intervento previste in uno studio clinico possono essere singole o combinazioni di farmaci diagnostici, preventivi (come i vaccini) o terapeutici, prodotti biologici, dispositivi, schemi terapeutici, procedure o approcci educativi. Le tecniche di intervento dovrebbero essere applicate ai partecipanti in modo standardizzato, con l'obiettivo di modificare un determinato esito. Il monitoraggio nel tempo dei soggetti, in assenza di un intervento attivo, può servire a descrivere la storia naturale di un processo patologico, ma non costituisce uno studio clinico. Senza un intervento attivo, lo studio è considerato osservazionale, poiché non viene condotto alcun esperimento. Essendo lo studio clinico un esperimento sugli esseri umani, gli sponsor e coloro che lo conducono hanno obblighi etici nei confronti dei partecipanti alla sperimentazione e nei confronti della scienza e della medicina. Infatti, gli effetti di un intervento non includono soltanto l'efficacia di uno specifico trattamento, ad esempio, ma anche la sua sicurezza. Allo stesso tempo il valore del trattamento tiene conto sia dell'utilità sociale (necessità individuali e collettive) di un intervento sia delle questioni etiche (rispetto dei diritti fondamentali).
Uno studio clinico è costituito da diverse fasi successive, che verranno analizzate più dettagliatamente nella sezione 1.5 di questo capitolo. Nelle fasi iniziali, gli studi possono essere controllati o non controllati. Sebbene la terminologia comune faccia riferimento a trial di fase I e fase II, poiché questi sono talvolta non controllati, li definiremo semplicemente studi clinici. Uno studio clinico controllato (trial), secondo la definizione data, include un gruppo di controllo con cui viene confrontato il gruppo che riceve l'intervento. Al momento iniziale dello studio (*baseline*), il gruppo di controllo deve essere sufficientemente simile al gruppo di intervento per quanto riguarda gli aspetti rilevanti, in modo che eventuali differenze negli esiti possano essere ragionevolmente attribuite all'effetto dell'intervento. Tra questi aspetti si possono considerare, ad esempio, le caratteristiche demografiche (età, sesso), lo stato di salute generale, la gravità della patologia in studio, la presenza di comorbidità oppure variabili biologiche e cliniche misurabili (ad es. pressione arteriosa, indice di massa corporea, valori di laboratorio). Tali caratteristiche, che possono influenzare la risposta all'intervento, sono dette *covariate*. L'adeguato bilanciamento delle covariate tra i gruppi rappresenta un prerequisito fondamentale per garantire la validità interna del trial.
Nella maggior parte dei casi, un nuovo trattamento viene confrontato con la migliore

terapia standard attualmente disponibile oppure viene usato in combinazione con essa. Solo nel caso in cui tale standard non esista oppure non sia disponibile per diversi motivi, è appropriato che i partecipanti del gruppo di intervento vengano confrontati con soggetti che non ricevono alcun trattamento attivo. "Nessun trattamento attivo" significa che il partecipante può ricevere un placebo oppure nessun trattamento. Ovviamente, i partecipanti di entrambi i gruppi possono essere sottoposti a terapie e regimi aggiuntivi, i cosiddetti trattamenti concomitanti, che possono essere auto-somministrati oppure prescritti da altri medici (ad esempio, da specialisti o medici di base).
Lo studio clinico ideale è uno studio *randomizzato* e *in doppio cieco*. Qualsiasi deviazione da questo standard comporta potenziali svantaggi, che saranno discussi principalmente nelle sezioni 1.2 e 1.4. In alcuni studi clinici, i compromessi sono inevitabili, ma spesso le carenze possono essere prevenute o ridotte al minimo applicando i principi fondamentali di progettazione (sezione 1.3), conduzione e analisi (capitoli 2 e 3). Gli studi clinici randomizzati, se ben progettati e condotti su campioni sufficientemente ampi, rappresentano il miglior metodo disponibile per stabilire quali interventi siano efficaci e generalmente sicuri, contribuendo così al miglioramento della salute pubblica. Purtroppo, solo una minoranza delle raccomandazioni contenute nelle linee guida cliniche si basa su evidenze derivanti da studi randomizzati, ovvero il tipo di evidenza necessario per poter avere fiducia nei risultati. Pertanto, sebbene i trial costituiscano la base essenziale dell'evidenza scientifica, molte terapie e misure preventive di uso comune non sono supportate da studi clinici randomizzati. Potenziare la capacità, la qualità e la rilevanza degli studi clinici rappresenta quindi una priorità fondamentale per la salute pubblica.

1.2 Processo di randomizzazione

Gli *studi clinici randomizzati e controllati* (RCT) sono studi comparativi che prevedono un gruppo di intervento e un gruppo di controllo per ognuno dei quali l'assegnazione dei partecipanti avviene tramite una procedura di allocazione casuale. La randomizzazione, nel caso più semplice, è un processo mediante il quale tutti i partecipanti hanno la stessa probabilità di essere assegnati al gruppo di intervento o al gruppo di controllo. L'allocazione randomizzata presenta tre vantaggi rispetto ad altri metodi di assegnazione dei pazienti ai diversi bracci dello studio [10].

1. Il primo vantaggio è che la randomizzazione elimina il rischio di bias (distorsione)

nell'assegnazione dei partecipanti al gruppo di intervento o al gruppo di controllo, evitando che il processo di allocazione riuslti prevedibile. Tale *bias di selezione* può insorgere facilmente (e non è necessariamente evitabile) negli studi con controlli storici o concorrenti non randomizzati, poiché l'assegnazione dei partecipanti ai due gruppi può essere influenzata da fattori prognostici o da decisioni soggettive del ricercatore o del partecipante. Ad esempio, un medico potrebbe decidere di somministrare il nuovo intervento solo ai pazienti con una prognosi migliore oppure i pazienti stessi potrebbero rifiutare il trattamento sperimentale preferendo la terapia standard. Tali decisioni, consapevoli o inconsapevoli, possono essere influenzate da diversi fattori, come la gravità della malattia, le preferenze personali o le aspettative sul trattamento e finiscono per creare gruppi non comparabili, compromettendo la validità del confronto. Questo vantaggio della randomizzazione presuppone che la procedura sia eseguita in modo valido e che l'assegnazione non sia prevedibile.

2. Il secondo vantaggio, correlato al primo, è che la randomizzazione tende a generare gruppi comparabili eliminando il cosidetto *bias accidentale*, che può emergere se la procedura di assegnazione non riesce a bilanciare adeguatamente i gruppi rispetto ai fattori di rischio o alle covariate prognostiche. Se la procedura è condotta correttamente, le caratteristiche prognostiche (sia note che ignote o non misurate) dei partecipanti, al momento della randomizzazione, saranno in media distribuite in modo equilibrato tra il gruppo di intervento e il gruppo di controllo. Ciò non significa che in un singolo esperimento tutte queste caratteristiche saranno perfettamente bilanciate tra i due gruppi. Significa però che, per covariate indipendenti, qualunque differenza (rilevata o non rilevata) sarà, in media, equamente distribuita in termini di entità e direzione tra i due gruppi. In altri termini, la randomizzazione non elimina del tutto le differenze tra i gruppi in un singolo esperimento, ma garantisce che eventuali squilibri siano casuali, non sistematici e statisticamente gestibili. È questa casualità che protegge dall'introduzione di bias e rende validi i confronti tra gruppi. Alcune delle tecniche di assegnazione sono più suscettibili a questo tipo di bias, in particolare negli studi di piccole dimensioni. Tuttavia, negli studi di grandi dimensioni, la probabilità che si verifichi un bias accidentale è trascurabile [23].

3. Il terzo vantaggio della randomizzazione consiste nel fatto che essa garantisce la validità dei test statistici di significatività (che saranno definiti più rigorosamente nel prossimo capitolo). Sebbene in un singolo esperimento i gruppi confrontati non risultino mai perfettamente bilanciati rispetto a tutte le covariate rilevanti, la randomizzazione consente di attribuire una distribuzione di probabilità alle differenze osservate negli esiti, assumendo che i trattamenti abbiano pari efficacia. In tal modo, è possibile assegnare un livello di significatività statistica alle differenze riscontrate tra i gruppi. In termini metodologici, la randomizzazione permette di valutare, in assenza di un effetto differenziale tra i trattamenti (ipotesi nulla), quanto sia plausibile che una differenza osservata tra gruppi sia attribuibile al caso. Pertanto, anche in presenza di squilibri casuali nelle covariate, l'applicazione di test statistici standard, come il *chi quadrato* o il *test t di Student*, rimane giustificata. Al contrario, negli studi non randomizzati è necessario formulare ipotesi aggiuntive sulla comparabilità dei gruppi (si deve assumere che i gruppi siano sufficientemente simili in termini di caratteristiche importanti) e sulla correttezza dei modelli statistici utilizzati (il modello statistico deve descrivere correttamente i dati) affinché i test di significatività risultino validi. Tuttavia, la verifica empirica di tali assunzioni è spesso complessa e incerta, con un conseguente rischio di bias nell'interpretazione dei risultati.

1.2.1 Randomizzazione a proporzione fissa

Le procedure di allocazione (assegnazione) casuale a proporzione fissa prevedono che gli interventi vengano assegnati ai partecipanti secondo una probabilità predefinita, solitamente uguale per tutti i gruppi e tale probabilità non viene modificata nel corso dello studio. L'assegnazione ai gruppi di intervento e controllo dovrebbe avvenire in modo equilibrato (rapporto di allocazione 1:1), salvo che non vi siano motivi validi per fare altrimenti. Un rapporto di assegnazione sbilanciato, ad esempio 2:1 a favore del gruppo di intervento, potrebbe essere utile in specifiche circostanze. La logica di tale scelta è che, sebbene lo studio possa perdere una minima parte di potenza statistica, si può ottenere maggiore informazione sulle risposte dei partecipanti al nuovo intervento, in particolare per quanto riguarda tossicità ed effetti collaterali. In alcuni casi, può essere sufficiente raccogliere meno dati sul gruppo di controllo e, quindi, coinvolgere un numero inferiore

di partecipanti in quel gruppo. Inoltre, se l'intervento si dimostra benefico, un maggior numero di partecipanti ne trarrebbe vantaggio rispetto a una procedura con allocazione equa. Tuttavia, se l'intervento si rivelasse dannoso, un numero maggiore di soggetti ne sarebbe esposto in presenza di una strategia di assegnazione sbilanciata. Benché la perdita di potenza associata a rapporti di allocazione compresi tra 1/2 e 2/3 sia in genere inferiore al 5% [3, 35], la procedura con allocazione equa rimane quello più potente e quindi generalmente raccomandata. Inoltre, l'allocazione con rapporto 1:1 risulta più coerente con il *principio di indifferenza* (*equipoise*), secondo cui non si presuppone una superiorità iniziale di uno dei due trattamenti. L'adozione di un'allocazione sbilanciata può infatti trasmettere ai partecipanti o ai loro medici curanti l'idea che uno dei due interventi sia preferibile all'altro. In alcune situazioni specifiche, l'elevato costo di un trattamento può giustificare un'allocazione 2:1 o 3:1 per contenere i costi, senza determinare una perdita significativa di potenza. In ogni caso, queste scelte comportano dei compromessi da valutare attentamente. Nel prosieguo della trattazione, si presumerà un'allocazione equa, salvo diversa indicazione.
Esistono diverse metodologie per assegnare casualmente i pazienti mantenendo una proporzione fissa tra il gruppo di intervento e quello di controllo [22, 24, 39, 53] e, in questa sede, ne verranno analizzate due: *randomizzazione semplice* e *randomizzazione a blocchi*.

1.2.1.1 Randomizzazione semplice

La randomizzazione semplice è la forma più elementare di allocazione a proporzione fissa e prevede che i pazienti vengano assegnati ai due bracci in maniera completamente casuale. Questa procedura equivale a lanciare una moneta non truccata ogni volta che un partecipante è eleggibile per la randomizzazione. Ad esempio, se esce testa, il partecipante viene assegnato al gruppo A; se esce croce, al gruppo B. In questo modo, circa la metà dei partecipanti sarà assegnata al gruppo A e l'altra metà al gruppo B.
Nella pratica, soprattutto negli studi di piccole dimensioni, al posto del lancio della moneta viene spesso utilizzata una tabella di cifre casuali, in cui le cifre da 0 a 9 (tutte con la stessa probabilità) sono disposte in righe e colonne. Selezionando casualmente una riga (o una colonna) e leggendo la sequenza di cifre, si può ad esempio assegnare il gruppo A ai partecipanti per cui la cifra successiva è pari e il gruppo B a quelli per cui è dispari. Questo processo genera una sequenza casuale di assegnazioni e ciascun partecipante ha la stessa probabilità di essere assegnato al gruppo A o al gruppo B.

Negli studi di grandi dimensioni, un metodo più pratico per generare una lista di randomizzazione consiste nell'utilizzare un algoritmo per la generazione di numeri casuali, disponibile nella maggior parte dei sistemi informatici. Una procedura di randomizzazione semplice può assegnare i partecipanti al gruppo A con probabilità p e al gruppo B con probabilità $1 - p$. Un metodo informatizzato per realizzare questa randomizzazione consiste nell'utilizzare un algoritmo di numeri casuali uniformemente distribuiti nell'intervallo tra 0 e 1. Per ogni partecipante viene generato un numero casuale: se il numero è compreso tra 0 e p, il partecipante viene assegnato al gruppo A; altrimenti, viene assegnato al gruppo B. Nel caso di allocazione equa, il valore di p sarà pari a 0.5. Se si desidera una allocazione sbilanciata ($p \neq 0.5$), allora p può essere impostato in base alla proporzione desiderata e lo studio avrà, in media, una proporzione p di partecipanti assegnati al gruppo A.
Il vantaggio principale della randomizzazione semplice è rappresentato dalla facilità di implementazione. Tuttavia, presenta anche uno svantaggio rilevante: sebbene nel lungo periodo il numero di partecipanti in ciascun gruppo tenda a rispettare la proporzione prestabilita, in qualsiasi momento del processo di assegnazione, incluso il termine dello studio, può verificarsi un significativo sbilanciamento. Questo rischio è particolarmente evidente negli studi con piccoli campioni. Sebbene questi squilibri non invalidino i test statistici, essi ridimensionano la capacità dello studio di rilevare reali differenze tra i due gruppi. Inoltre, squilibri visibili nei numeri possono apparire problematici dal punto di vista della presentazione e minare la credibilità dello studio, soprattutto per chi non ha una formazione statistica. Per queste ragioni, la randomizzazione semplice non viene spesso utilizzata, neppure negli studi di grandi dimensioni.

1.2.1.2 Randomizzazione a blocchi

Nella randomizzazione a blocchi si suddivide la sequenza di pazienti eleggibili in blocchi di dimensione pari prefissata (ad esempio 4, 6 o 8) e, all'interno di ciascun blocco, le assegnazioni ai gruppi vengono generate in modo casuale, ma rispettando la proporzione di allocazione predefinita (ad esempio 1:1 oppure 2:1). In tal modo, alla fine di ogni blocco, i due gruppi hanno un numero di pazienti che riflette esattamente il rapporto di allocazione desiderato. Il processo viene ripetuto per blocchi consecutivi di partecipanti, fino a completare l'assegnazione per l'intero campione.

Ad esempio, i ricercatori potrebbero voler garantire che, ogni quattro partecipanti randomizzati, il numero di soggetti nei due bracci dello studio sia uguale (rapporto di allocazione 1:1). In tal caso, si utilizzerebbe un blocco di dimensione 4 e il processo consisterebbe nel randomizzare l'ordine di assegnazione di due partecipanti al gruppo A e due al gruppo B per ogni sequenza di quattro soggetti reclutati nello studio. È possibile elencare tutte le combinazioni possibili di assegnazione all'interno di un blocco e poi randomizzare l'ordine con cui queste combinazioni vengono selezionate. Nel caso di un blocco di dimensione 4, esistono sei combinazioni possibili per assegnare due A e due B: $AABB, ABAB, BAAB, BABA, BBAA, ABBA$. Una di queste sequenze viene selezionata casualmente e i quattro partecipanti successivi vengono assegnati secondo quell'ordine. Questo processo viene poi ripetuto tutte le volte necessarie, fino a completare l'assegnazione per l'intera popolazione dello studio.
La randomizzazione a blocchi consente di evitare squilibri rilevanti nel numero di partecipanti assegnati a ciascun gruppo, squilibri che potrebbero invece verificarsi con la procedura di randomizzazione semplice. Più precisamente, la randomizzazione a blocchi garantisce che, durante tutto il processo di reclutamento, lo sbilanciamento tra i gruppi non si discosti eccessivamente dal rapporto di allocazione prefissato e che, in determinati momenti del reclutamento (alla fine di ogni blocco), il numero di partecipanti nei bracci rispecchi esattamente tale rapporto. Questo metodo rappresenta una protezione efficace contro eventuali tendenze temporali durante la fase di arruolamento, una criticità frequente nei trial di grandi dimensioni con fasi di inclusione prolungate nel tempo.
In particolare, se il rapporto di allocazione è 1:1, la differenza nel numero di soggetti tra i gruppi non supererà mai $b/2$, dove b rappresenta la dimensione del blocco. Questo aspetto può essere importante per almeno due ragioni. In primo luogo, se il profilo dei partecipanti reclutati cambia nel corso del periodo di inclusione oppure la malattia in esame ha un'incidenza variabile nel tempo, la randomizzazione a blocchi contribuisce a ottenere gruppi più comparabili. Ad esempio, un ricercatore potrebbe utilizzare fonti diverse di reclutamento in momenti successivi. I partecipanti provenienti da queste fonti potrebbero differire in termini di gravità della malattia o per altre caratteristiche rilevanti. Una fonte, che include soggetti più gravi, potrebbe essere utilizzata all'inizio dell'arruolamento, mentre un'altra, con soggetti più sani, verso la fine. Se la randomizzazione non fosse a blocchi, è possibile che un numero maggiore di pazienti gravemente malati venga assegnato a uno dei due gruppi. Dato che i partecipanti successivi sono

meno gravi, questo squilibrio iniziale non verrebbe corretto nel tempo.
In secondo luogo, un ulteriore vantaggio della randomizzazione a blocchi è rappresentato dal fatto che, nel caso in cui lo studio venga interrotto prima del completamento del reclutamento, verrà comunque mantenuto un equilibrio numerico tra i partecipanti assegnati ai vari gruppi.
Dal punto di vista teorico, la randomizzazione a blocchi presenta il limite di rendere più complessa l'analisi statistica dei dati rispetto alla randomizzazione semplice. Se l'analisi finale dei dati non riflette correttamente il tipo di randomizzazione effettivamente utilizzata, può produrre risultati errati, poiché i metodi analitici standard presuppongono generalmente l'uso della randomizzazione semplice. Nella pratica, molti ricercatori ignorano il fatto che la randomizzazione sia stata effettuata a blocchi durante l'analisi. Matts e McHugh [32] hanno studiato questo problema, concludendo che la stima della variabilità utilizzata nell'analisi statistica non è esattamente corretta se la struttura a blocchi viene trascurata. Di norma, l'analisi che ignora i blocchi è conservativa [22]. In altre parole, un'analisi che non tiene conto dei blocchi potrebbe avere una potenza leggermente inferiore rispetto a quella corretta e sottostimare il reale livello di significatività statistica. Di conseguenza, la randomizzazione a blocchi accompagnata da un'analisi statistica appropriata è generalmente più potente rispetto a non usare affatto i blocchi oppure a utilizzarli ma ignorarli nell'analisi. Tuttavia, trattare correttamente la presenza di blocchi può risultare difficile da estendere ad analisi più complesse, come quelle che coinvolgono covariate, sottogruppi o analisi secondarie. La possibilità di utilizzare un approccio analitico unico e diretto, in grado di gestire questi aspetti senza complicazioni aggiuntive, semplifica l'interpretazione complessiva dello studio.

1.3 Progettazione dei trial clinici

La progettazione di uno studio clinico si fonda sulla definizione preliminare della domanda di ricerca, da cui derivano sia la struttura dello studio sia le analisi statistiche da condurre. In particolare, è fondamentale distinguere tra domanda primaria, sulla quale si fonda l'intero piano sperimentale, e domande secondarie, che arricchiscono l'analisi ma non devono compromettere l'interpretabilità dei risultati.

1.3.1 Domande di ricerca

La domanda primaria è l'interrogativo principale cui il trial intende rispondere ed è strettamente collegata alla variabile di risposta primaria. Essa deve essere chiaramente definita prima dell'inizio dello studio, clinicamente rilevante e statisticamente testabile. Inoltre, sulla base della domanda primaria viene calcolata la dimensione campionaria ed è questa l'analisi che determina l'esito principale dello studio.
Nel contesto di un trial vaccinale, ad esempio, la domanda primaria consiste nel chiedersi se il vaccino sperimentale riduce l'incidenza della malattia rispetto al placebo nei soggetti adulti non precedentemente infettati.
Le domande secondarie rappresentano interrogativi aggiuntivi che possono riguardare altri esiti clinici rilevanti (es. ospedalizzazione, mortalità, eventi avversi), sottogruppi di popolazione (es. fasce di età, comorbidità, varianti virali) e variabili esploratorie. Benché utili per comprendere meglio l'efficacia e la sicurezza di un intervento, tali domande devono essere fondate su ipotesi plausibili e limitate nel numero per evitare un tasso di falsi positivi elevato (errore statistico di tipo I, definito nel capitolo 2).
Un esempio di domanda secondaria in un trial vaccinale potrebbe consistere nel chiedersi se il vaccino in esame riduce anche la probabilità di ospedalizzazione per forme gravi della malattia.

1.3.2 Variabili di risposta e obiettivi di efficacia

Nei trial clinici la *variabile di risposta* rappresenta la quantità osservabile associata a ciascun partecipante, cioè il dato grezzo raccolto e registrato, che risponde alle domande di ricerca poste. Essa può essere di natura dicotomica (es. malattia presente/assente), quantitativa continua (es. valori pressori, livelli ematici) oppure di tipo *time-to-event* (tempo alla comparsa di un evento). In quest'ultimo caso, come si vedrà nella seconda parte, la variabile di risposta formale non è semplicemente l'indicatore dell'evento, ma una coppia aleatoria che tiene conto anche della possibile censura, ovvero il tempo massimo per cui un partecipante, dopo essere stato reclutato, viene seguito fino alla fine dello studio. In questo modo ogni soggetto contribuisce con l'informazione disponibile, anche se non è andato incontro all'evento durante l'osservazione.
Dal punto di vista clinico, le variabili di risposta possono essere distinte in variabili dirette (es. decesso, infezione confermata, ricovero), variabili soggettive (es. miglioramento dei sintomi, qualità della vita) e variabili surrogate o biologiche (es. carica virale, titolo

anticorpale, PCR negativa). Una variabile surrogata è un indicatore sostitutivo, misurabile più facilmente e rapidamente , che viene usato negli studi clinici al posto di un vero esito clinico per stimarne indirettamente l'efficacia. L'uso di variabili surrogate è particolarmente diffuso nelle fasi iniziali di uno studio (fase I–II) per valutare l'impatto biologico del trattamento, ma non è sostitutivo delle variabili cliniche dirette nei trial di efficacia confermativi (fase III), dove è necessario dimostrare un beneficio tangibile per i pazienti.
Accanto alla variabile di risposta è necessario distinguere il concetto di *obiettivo di efficacia* (*endpoint*). L'endpoint non è una variabile osservabile sul singolo individuo, ma un criterio analitico predefinito su cui si fonda la valutazione principale del trial. In altre parole, l'endpoint è l'elaborazione statistica delle variabili di risposta raccolte nei vari gruppi di studio, con lo scopo di stimare l'efficacia o la sicurezza dell'intervento.
Una corretta formulazione della domanda primaria e la selezione rigorosa della variabile di risposta sono elementi essenziali per garantire validità interna, rilevanza clinica e credibilità dei risultati di uno studio. Le domande secondarie e le variabili esploratorie devono essere gestite con cautela, per evitare inferenze fuorvianti dovute a molteplicità di test o ad analisi post-hoc (cioè analisi non pianificate ed effettuate solo dopo aver osservato i dati).

1.3.3 Disposizione dei gruppi di trattamento

Negli studi clinici si adottano diversi tipi di *design* sperimentali, che si possono classificare in base alla disposizione dei gruppi di trattamento oppure in base alla flessibilità e all'adattamento durante lo studio.
La maggior parte degli studi clinici adotta il cosiddetto *design parallelo*, in cui i gruppi di intervento e di controllo vengono osservati in modo simultaneo a partire dal momento dell'assegnazione. In questo schema, ciascun partecipante resta assegnato al proprio gruppo per l'intera durata dello studio e il confronto tra i gruppi avviene lungo un medesimo arco temporale. Una variante del design parallelo è rappresentata dal *design cross-over*, nel quale ciascun partecipante riceve entrambi i trattamenti (intervento e controllo) in momenti distinti. In questo tipo di piano sperimentale, ogni soggetto contribuisce ai dati di entrambi i gruppi, permettendo un confronto intra-soggetto. Questo consente di ridurre la variabilità associata alle differenze individuali tra partecipanti poiché l'effetto dell'intervento viene misurato come la differenza nella risposta individuale del

partecipante tra la somministrazione del trattamento e quella del controllo. Tuttavia, l'applicazione del design cross-over richiede condizioni metodologiche specifiche, come l'assenza di effetti permanenti del trattamento e la previsione di un adeguato periodo di *washout* tra le fasi di trattamento, al fine di evitare effetti residui.

1.3.4 Design adattivi

Esiste un grande interesse verso i cosiddetti *design adattivi*, anche se con questa definizione si indicano tipologie diverse di design sperimentale, che fanno riferimento a concetti differenti di "adattività". In senso classico, l'adattività si riferisce alla possibilità di modificare il protocollo sperimentale in corso d'opera sulla base di informazioni accumulate durante lo studio, nel rispetto di predefiniti vincoli metodologici e statistici. Tali modifiche possono includere, ad esempio, l'aggiustamento delle dosi, la modifica della randomizzazione, l'aggiunta o eliminazione di bracci di trattamento o il restringimento a sottogruppi di pazienti in funzione della risposta osservata.
Altri design sono considerati adattivi sul piano temporale e campionario, in risposta ai dati che si raccolgono. Per esempio, un design *event-driven* è una strategia di progettazione sperimentale in cui la durata dello studio o il numero di partecipanti non sono prefissati rigidamente, bensì determinati in funzione del numero di eventi osservati (ad esempio, infezioni, decessi, recidive), considerato sufficiente per garantire una potenza statistica adeguata. In questo contesto, si parla di adattività non in riferimento alla modifica dei trattamenti in risposta ai dati intermedi, bensì in relazione alla flessibilità temporale e campionaria dello studio. A differenza di un design *time-driven*, in cui la durata dello studio è fissata a priori, in un design event-driven la raccolta dei dati prosegue fino al raggiungimento di una soglia predefinita di eventi, indipendentemente dalla durata prevista o dal numero finale di soggetti reclutati. Questo approccio consente di adattare la durata dello studio all'incidenza reale dell'evento, risultando più efficiente e robusto in presenza di incertezza sulla frequenza attesa degli esiti. In caso di eventi più rari del previsto, lo studio può estendersi nel tempo o aumentare il numero di soggetti arruolati; al contrario, se gli eventi si accumulano rapidamente, può concludersi prima, ottimizzando tempi e risorse. Tale flessibilità rende il design event-driven particolarmente adatto in ambito clinico quando l'esito di interesse è un evento binario o un tempo all'evento (time to event).

1.3.5 Protocollo di studio

Ogni trial clinico ben progettato richiede un protocollo. Il protocollo di studio può essere considerato come un accordo scritto tra il ricercatore, il partecipante e la comunità scientifica. Esso fornisce le informazioni di background, specifica gli obiettivi e descrive il design e l'organizzazione dello studio. Non è necessario includere ogni singolo dettaglio operativo, a condizione che tali informazioni siano contenute in un manuale operativo o manuale delle procedure a supporto del protocollo. Il protocollo svolge anche un'importante funzione di comunicazione tra tutti coloro che collaborano allo studio e dovrebbe essere disponibile pubblicamente su richiesta. Molti protocolli, oggi, vengono pubblicati su riviste scientifiche online. Il protocollo dovrebbe essere redatto prima dell'inizio del reclutamento dei partecipanti e dovrebbe rimanere sostanzialmente invariato nel tempo, salvo modifiche minori. Eventuali modifiche devono essere attentamente valutate e giustificate. Le revisione sostanziali, che alterano la direzione dello studio, dovrebbero essere rare e, se effettuate, è fondamentale documentare chiaramente la motivazione e il processo decisionale che le ha determinate.

1.4 Il mascheramento negli studi clinici

Il *mascheramento* (*blinding*) rappresenta una delle strategie fondamentali nella progettazione degli studi clinici controllati, con l'obiettivo di minimizzare i bias sistematici che possono compromettere la validità interna dei risultati. Attraverso il mascheramento si limita la possibilità che le aspettative, le preferenze o i comportamenti dei partecipanti e degli sperimentatori influenzino i dati raccolti o l'interpretazione degli esiti.

1.4.1 Tipologie di bias prevenibili

Il mascheramento è principalmente volto a evitare tre tipologie di bias [21].

1. Il *bias di performance* si verifica quando la consapevolezza dell'intervento ricevuto influenza il comportamento dei partecipanti o dei medici. Ad esempio, i pazienti che sanno di ricevere il trattamento attivo potrebbero modificare le proprie abitudini in modo diverso rispetto a quelli assegnati al gruppo di controllo, mentre i medici potrebbero fornire cure aggiuntive o più attente.

2. Il *bias di rilevazione* (*detection bias*) si manifesta quando il valutatore degli esiti, sapendo quale trattamento è stato assegnato, interpreta o misura i risultati in modo distorto, soprattutto nel caso di esiti soggettivi.

3. Infine, il *bias da abbandono* (*attrition bias*) riguarda la probabilità di abbandono selettivo dello studio, che può aumentare nei partecipanti che percepiscono di non ricevere il trattamento desiderato, alterando l'equilibrio della randomizzazione.

1.4.2 Forme di mascheramento

Le forme principali di mascheramento si distinguono in base a quali figure sono mantenute all'oscuro dell'assegnazione. Nella modalità *non in cieco* (*open-label*), né i partecipanti né il personale di studio sono mascherati: ciò comporta il massimo rischio di bias, ma talvolta è inevitabile, ad esempio negli studi su procedure chirurgiche o interventi comportamentali. Lo studio in *cieco singolo* (*single-blind*) prevede che solo i partecipanti non conoscano l'intervento ricevuto: questo riduce il bias di reporting da parte del paziente, ma non protegge dalla possibilità che lo sperimentatore influenzi la gestione clinica o la raccolta dati. Lo studio in *doppio cieco* (*double-blind*) rappresenta lo standard metodologico ideale: né i partecipanti né il team clinico e di raccolta dati conoscono l'assegnazione, con una riduzione significativa del rischio di bias di performance e di rilevazione. Infine, nello studio *triplo cieco* (triple-blind) anche il comitato di monitoraggio dei dati rimane inizialmente all'oscuro della distribuzione dei trattamenti. Questa modalità, adottata in casi particolari, può migliorare l'oggettività nell'interpretazione intermedia dei risultati, ma comporta rischi etici se ostacola la tempestiva individuazione di segnali di danno.
La rimozione del mascheramento (unblinding), intenzionale o accidentale, può compromettere l'integrità dello studio. Sebbene in alcuni casi clinici sia necessaria (es. in caso di gravi eventi avversi), andrebbe evitata il più possibile, poiché espone nuovamente lo studio al rischio dei bias sopra descritti. La perdita del cieco può portare a una modifica nel comportamento del paziente o del personale, influenzando l'aderenza, la prescrizione di terapie concomitanti o la valutazione degli esiti. Per contenere questi effetti, è buona prassi documentare ogni episodio di unblinding, limitarlo al personale non coinvolto nella valutazione degli esiti e includere analisi di sensibilità per valutare il potenziale impatto. L'applicazione rigorosa del mascheramento, quando tecnicamente possibile, costituisce una misura essenziale per tutelare la validità interna dello studio clinico. Essa consente

di isolare l'effetto del trattamento in esame da interferenze cognitive e comportamentali, garantendo che le differenze osservate tra i gruppi siano attribuibili esclusivamente all'intervento sperimentale. Anche se non sempre realizzabile (soprattutto in studi non farmacologici), il mascheramento rimane uno degli strumenti più efficaci per prevenire bias e rafforzare la robustezza delle evidenze prodotte.

1.5 Le fasi di uno studio clinico

Lo sviluppo di nuovi interventi clinici si articola in quattro fasi principali (Fase I–IV), ciascuna delle quali risponde a obiettivi distinti e richiede strategie di design e analisi differenziate. Dal punto di vista statistico, ogni fase pone specifici problemi di inferenza, pianificazione campionaria e modellazione dei dati.

1.5.1 Fase I – Studi esplorativi di sicurezza e farmacologia

Gli studi di fase I rappresentano la prima applicazione nell'uomo del trattamento in esame. L'obiettivo primario è stimare la sicurezza e la tollerabilità della nuova terapia, nonché descrivere le sue caratteristiche farmacocinetiche e farmacodinamiche.
Dal punto di vista statistico, questi studi adottano frequentemente design a *incremento graduale della dose* (*dose-escalation*), con assegnazione sequenziale di piccoli gruppi di pazienti a dosi crescenti. L'obiettivo è identificare la *dose massima tollerata* (MTD), ossia la dose alla quale la probabilità di tossicità supera una soglia predefinita.
Ad esempio, nei primi studi sui vaccini anti-COVID-19 si è adottata una logica esplorativa per valutare la sicurezza di diverse formulazioni e dosaggi. I dati raccolti includevano eventi avversi, livelli anticorpali e parametri farmacocinetici.

1.5.2 Fase II – Stima preliminare dell'efficacia

La fase II mira a stimare in modo preliminare la probabilità di successo del trattamento. Gli obiettivi includono la valutazione dell'attività biologica, la caratterizzazione della dose-risposta e la *selezione del dosaggio ottimale* per la fase III. Statisticamente, questa fase è spesso supportata da *design a due stadi* (come i design di Gehan [15] o Simon [48]), con l'obiettivo di escludere trattamenti inefficaci e minimizzare l'esposizione non necessaria. Possono essere utilizzati anche *design bayesiani* [6], che permettono l'aggiornamento continuo delle probabilità di risposta e l'adattamento del campione. Le

inferenze sono generalmente centrate su endpoint surrogati (cioè criteri di valutazione basati su variabili di risposta surrogate) o biomarcatori.
Nella fase II dei vaccini COVID-19 sono state testate più dosi per valutare la risposta immunitaria e selezionare la formulazione con il miglior compromesso tra efficacia immunogenica e sicurezza.

1.5.3 Fase III – Studio controllato confermativo

Gli studi di fase III costituiscono il fulcro inferenziale della sperimentazione clinica. Si tratta di RCT su larga scala, progettati per fornire evidenza conclusiva di efficacia clinica, confrontando l'intervento con un controllo attivo o placebo.
Statisticamente, la fase III prevede l'utilizzo di un endpoint primario predefinito e validato, la formulazione esplicita di ipotesi, il calcolo della potenza statistica e l'adozione di metodi rigorosi di analisi. Inoltre, in questa fase, si applicano spesso metodi di analisi ad interim, regole di stopping e correzioni per test multipli [2, 7, 33, 49].
Gli RCT dei vaccini COVID-19 (es. Pfizer/BioNTech, Moderna) hanno randomizzato decine di migliaia di partecipanti per confrontare l'incidenza cumulativa di casi sintomatici tra bracci vaccino e placebo. Sono stati utilizzati modelli di regressione del rischio [8, 17], test di log-rank [36, 40] e intervalli di confidenza e di credibilità per stimare l'efficacia vaccinale. Questi intervalli statistici verranno trattati, in termini generali, nel Capitolo 2 e, più nel dettaglio, nel Capitolo 3, mentre per un'introduzione ai metodi di analisi ad interim, alle regole di stopping, alle correzioni per test multipli, ai modelli di regressione e ai test di log-rank si rimanda ai riferimenti sopraindicati. Inoltre, nel Capitolo 4 gli intervalli statistici proposti saranno applicati allo studio clinico condotto da Pfizer/BioNTech tra il 2020 e il 2021 sul vaccino anti-COVID-19.

1.5.4 Fase IV – Sorveglianza post-registrativa e analisi di sicurezza

Gli studi di fase IV sono condotti dopo l'approvazione regolatoria del trattamento. L'obiettivo è valutare l'efficacia e la sicurezza nel contesto reale ("real-world evidence"), includendo popolazioni più eterogenee e orizzonti temporali estesi.
Dal punto di vista statistico, questa fase si basa prevalentemente su studi osservazionali, analizzati mediante modelli di rischio proporzionale [8], metodi del punteggio di propensione (propensity score methods [42]) o analisi bayesiane con prior informativi [20]. È fondamentale il controllo dei bias di selezione, confondimento, e eventi rari.

Ad esempio, dopo l'introduzione dei vaccini anti-COVID-19, studi osservazionali su milioni di individui hanno identificato effetti avversi rari (es. miocardite, trombosi) e stimato l'efficacia nel prevenire ospedalizzazioni e decessi, anche in presenza di nuove varianti virali.

Capitolo 2

Metodi statistici

2.1 Introduzione alla statistica

Questo secondo capitolo fornisce un'analisi matematica dettagliata dei principali metodi statistici (frequentisti e bayesiani) presenti in letteratura. Trattandosi di risultati classici e ampiamente consolidati, per le dimostrazioni dei teoremi enunciati si rimanda a [4], [43] e [13]. Nel prosieguo della trattazione si adotterà la seguente convenzione: i vettori saranno denotati mediante lettere in grassetto, gli scalari con lettere in carattere normale, le variabili aleatorie con lettere maiuscole e le corrispondenti realizzazioni con lettere minuscole.
La statistica è la scienza che si occupa di trarre conclusioni dai dati sperimentali. Una situazione tipica, con la quale bisogna spesso confrontarsi in ambito clinico o tecnologico, è quella in cui si studia un insieme molto grande di individui o di oggetti, detto *popolazione*, a cui sono associate delle quantità misurabili. L'approccio statistico consiste nell'estrarre un *campione rappresentativo* dalla popolazione di interesse, cioè un sottoinsieme che la rappresenti nelle sue caratteristiche fondamentali, e analizzarlo cercando di trarre delle conclusioni sulla popolazione nel suo insieme.
Per basare sui dati del campione delle inferenze che riguardino l'intera popolazione, è necessario assumere qualche condizione sulle relazioni che legano questi due insiemi. Un'ipotesi fondamentale, in molti casi del tutto ragionevole, consiste nell'assumere che le quantità misurabili associate agli oggetti estratti casualmente dalla popolazione siano variabili aleatorie indipendenti aventi la medesima distribuzione (densità) di probabilità. Infatti, se il campione viene selezionato in maniera completamente casuale, sembra

ragionevole supporre che i dati campionari relativi alla medesima grandezza siano valori identicamente distribuiti e indipendenti rispetto al processo di estrazione. La seguente definizione formalizza matematicamente il metodo di campionamento casuale per la raccolta dei dati.

Definizione 2.1. *Le variabili aleatorie $Y_1, \ldots, Y_n$ si dicono* ***campione aleatorio*** *(o* ***casuale****) di dimensione n dalla popolazione con densità di probabilità marginale $f(y)$ se $Y_1, \ldots, Y_n$ sono variabili aleatorie indipendenti e identicamente distribuite (i.i.d.) con densità $f(y)$.*

La distribuzione sottostante, identificata dalla densità $f(y)$, non è mai completamente nota, però si può sfruttare l'informazione contenuta nei dati campionari per fare inferenza su di essa secondo diversi approcci:

- se la distribuzione appartiene a una famiglia nota a meno di un insieme finito $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ di parametri incogniti, si parla di *inferenza parametrica*;
- se la distribuzione non ha una forma funzionale specifica (tranne al più assumere che sia continua o discreta), si parla di *inferenza non parametrica*;
- se la distribuzione è descritta da un numero finito di parametri $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ di interesse e, al tempo stesso, da componenti infinito-dimensionali (tipicamente funzioni di forma arbitraria) trattate come parametri di disturbo, si parla di *inferenza semiparametrica.*

Nel prosieguo della trattazione ci si focalizzerà unicamente sull'inferenza parametrica. Sebbene il valore di $\boldsymbol{\theta}$ non sia noto, si assume di conoscere l'insieme Θ dei suoi possibili valori, il cosiddetto *spazio parametrico.*
Denotando con $f(y \mid \boldsymbol{\theta})$ la densità di probabilità marginale di una distribuzione parametrica, la densità di probabilità congiunta di $Y_1, \ldots, Y_n$ è data da

$$f(y_1, \ldots, y_n \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\theta}),$$

dove $y_1, \ldots, y_n$ sono i valori osservati del campione, ognuno dei quali può variare nell'insieme dei punti $\mathcal{Y}$ in cui la densità $f(y \mid \boldsymbol{\theta})$, come funzione di y, è diversa da zero. Tale insieme è denominato *spazio campionario.*
Un campione casuale può essere interpretato come un vettore aleatorio, denotato $\mathbf{Y} =$

$(Y_1, ..., Y_n)$. A loro volta, le variabile aleatorie che compongono il campione possono essere variabili aleatorie multivariate, ovvero vettori aleatori costituiti da più grandezze misurabili. Un tipico esempio è fornito da dati time to event censurati, che costituiscono l'ipotesi fondamentale alla base del modello bayesiano proposto nel Capitolo 3. Per evitare confusione, le variabili che formano il campione sono denotate con la notazione scalare anche quando sono grandezze multidimensionali.

Esempio 2.1. *Si consideri un campione casuale di dimensione n in cui, per ciascun individuo i, con $i = 1, \dots, n$, si osservano dati di tipo time-to-event soggetti a censura e si assuma che l'evento di interesse (es. l'insorgenza di una malattia) si verificherebbe sempre se l'osservazione durasse sufficientemente a lungo. Per ogni individuo i, sia T_i la variabile aleatoria che rappresenta il tempo all'evento e sia C_i il tempo di censura non informativa, ossia una variabile aleatoria indipendente da T_i che riflette la fine del periodo di osservazione dell'individuo. Poiché, a causa della censura, non sempre l'evento viene osservato, i dati campionari risultano nella coppia*

$$Y_i = \big(\min(T_i, C_i),\ \mathbf{1}_{\{T_i < C_i\}} \big),$$

dove $\mathbf{1}_{\{T_i<C_i\}}$ è un'indicatrice che assume valore 1 se l'evento è stato osservato e 0 in caso contrario (censura), mentre $\min(T_i, C_i)$ è il tempo di sorveglianza dell'individuo i. Le realizzazioni $y_i = (s_i,\ \delta_i)$ costituiscono pertanto il campione osservato, il quale fornisce sia i tempi registrati sia l'informazione sulla presenza o meno dell'evento per ciascun individuo. Se i tempi all'evento sono esponenziali di parametro λ, denotando con $f_T(\cdot)$ la densità di T e con $f_C(\cdot)$ la densità di C, la densità congiunta del campione casuale è data da

$$\begin{aligned} f(y_1, \dots, y_n \mid \lambda) &= \underbrace{\prod_{i=1}^{n} [f_T(s_i)]^{\delta_i} [\bar{F}_T(s_i)]^{1-\delta_i}}_{\text{termini di } T} \times \underbrace{\prod_{i=1}^{n} [f_C(s_i)]^{1-\delta_i} [\bar{F}_C(s_i)]^{\delta_i}}_{\text{termini di } C} \\ &= \lambda^x e^{-\lambda s} \times \underbrace{\prod_{i=1}^{n} [f_C(s_i)]^{1-\delta_i} [\bar{F}_C(s_i)]^{\delta_i}}_{\text{termini non dipendenti da } \lambda}, \end{aligned} \tag{2.1}$$

dove $\bar{F}_T = \mathbb{P}(T > t)$ e $\bar{F}_C = \mathbb{P}(C > t)$ sono le funzioni di sopravvivenza di T e C rispettivamente, s è la somma dei tempi di sorveglianza e x è il numero di eventi osservati.

L'informazione contenuta nel campione $\mathbf{Y} = (Y_1, ..., Y_n)$ si può utilizzare per fare inferenza sul parametro (o sui parametri) θ. Quando la dimensione del campione n è elevata, l'insieme dei dati osservati $y_1, ..., y_n$ costituisce una lunga sequenza di numeri che può risultare di difficile interpretazione. Per questo motivo, è conveniente riassumere le informazioni contenute nel campione individuando alcune caratteristiche essenziali dei valori osservati. Tale obiettivo viene solitamente raggiunto attraverso il calcolo di statistiche, ossia funzioni del campione stesso.

Definizione 2.2. *Sia* $\mathbf{Y} = (Y_1, \ldots, Y_n)$ *un campione aleatorio di dimensione n estratto da una popolazione e sia* $T(y_1, \ldots, y_n)$ *una funzione a valori reali o vettoriali, il cui dominio include lo spazio campionario di* $\mathbf{Y}$*. Allora la variabile aleatoria, o il vettore aleatorio,*

$$Z = T(\mathbf{Y})$$

si dice ***statistica****. La distribuzione di probabilità di una statistica Z prende il nome di* ***distribuzione campionaria*** *di Z.*

Si osservi che, sebbene una statistica sia, per definizione, una funzione del solo campione, la sua distribuzione campionaria, in generale, dipende anche dal parametro (o dai parametri) θ della popolazione da cui il campione viene estratto. Infatti, qualsiasi statistica $T(\mathbf{Y})$ definisce una forma di riduzione o sintesi dei dati che preserva un'informazione almeno parziale su θ. Se si utilizza soltanto il valore osservato della statistica, $T(\mathbf{y})$, invece dell'intero campione osservato $\mathbf{y}$, due campioni $\mathbf{y}$ e $\mathbf{x}$ che soddisfano la condizione $T(\mathbf{y}) = T(\mathbf{x})$ risulteranno equivalenti, anche se i valori campionari effettivi possono differire sotto altri aspetti. Particolarmente utili sono le statistiche che permettono una riduzione dei dati senza perdere informazioni sul parametro incognito θ.

Definizione 2.3. *Una statistica* $T(\mathbf{Y})$ *si dice* ***sufficiente*** *per* θ *se la distribuzione condizionata del campione* $\mathbf{Y}$*, dato il valore di* $T(\mathbf{Y})$*, non dipende da* θ*.*

Una statistica sufficiente per un parametro θ è una statistica che, in un certo senso, cattura tutte le informazioni su θ contenute nel campione. Qualsiasi informazione addizionale presente nel campione, oltre al valore della statistica sufficiente, non apporta ulteriori informazioni riguardo a θ. Il seguente teorema, dovuto a Halmos e Savage (1949), consente di individuare una statistica sufficiente tramite una semplice ispezione della densità congiunta del campione.

Teorema 2.1 (Teorema di fattorizzazione)**.** *Sia $f(\mathbf{y} \mid \theta)$ la densità congiunta di un campione $\mathbf{Y}$. Una statistica $T(\mathbf{Y})$ è sufficiente per θ se e solo se esistono due funzioni $g(T(\mathbf{y}) \mid \theta)$ e $h(\mathbf{y})$ tali che, per ogni punto campionario $\mathbf{y}$ e per ogni valore del parametro θ,*

$$f(\mathbf{y} \mid \theta) = g(T(\mathbf{y}) \mid \theta)\, h(\mathbf{y}).$$

L'esempio 2.1 fornisce una semplice applicazione del teorema 2.1. Infatti, il primo fattore della distribuzione congiunta nell'equazione 2.1 dipende da λ e dai valori osservati delle statistiche

$$S = \sum_{i=1}^{n} \min(T_i, C_i) \quad \text{e} \quad X = \sum_{i=1}^{n} \mathbb{1}_{\{T_i < C_i\}},$$

mentre il secondo fattore dipende solo dal campione osservato $\mathbf{y}$. Dunque, la statistica vettoriale (S, X) è sufficiente per il parametro λ.

2.2 Distribuzioni campionarie asintotiche

Si rivela molto utile studiare il comportamento di specifiche statistiche campionarie quando la dimensione del campione tende all'infinito. Sebbene il concetto di dimensione campionaria infinita costituisca un artificio teorico, esso può rivelarsi utile per ottenere approssimazioni nel caso di campioni finiti, poiché, al tendere all'infinito della numerosità campionaria, le espressioni assumono spesso forme più semplici. In particolare, è degno di nota il comportamento asintotico della media campionaria

$$\overline{Y}_n = \frac{Y_1 + \cdots + Y_n}{n} = \frac{1}{n} \sum_{i=1}^{n} Y_i,$$

per n che tende all'infinito.
Data una successione di variabili aleatorie $\{Y_1, Y_2, \ldots\}$, esistono diverse nozioni di convergenza. Tuttavia, in questo contesto, si introduce soltanto una forma di convergenza in senso distribuzionale, che sarà utile nei prossimi capitoli.

Definizione 2.4. *Sia $\{Y_1, Y_2, \ldots\}$ una successione di variabili aleatorie e sia Y una variabile aleatoria. Si dice che Y_n* converge in distribuzione *a Y se, per ogni punto y in cui la funzione di ripartizione di Y è continua, vale*

$$\lim_{n \to \infty} F_{Y_n}(y) = F_Y(y),$$

dove:

- $F_{Y_n}(y) = \mathbb{P}(Y_n \leq y)$ *denota la funzione di ripartizione della variabile aleatoria* Y_n*;*
- $F_Y(y) = \mathbb{P}(Y \leq y)$ *denota la funzione di ripartizione della variabile limite* Y*.*

La convergenza in distribuzione si indica anche con il simbolo

$$Y_n \xrightarrow[n\to\infty]{d} Y.$$

In altri termini, al crescere di n, le distribuzioni delle Y_n si avvicinano, punto per punto nei punti di continuità di F_Y, alla distribuzione della variabile limite Y.

2.2.1 Teorema limite centrale

La distribuzione limite della media campionaria è riassunta in uno dei teoremi più significativi della statistica: il *Teorema limite centrale* (CLT).

Teorema 2.2 (Teorema limite centrale)**.** *Sia* $\{Y_1, Y_2, \ldots\}$ *una successione di variabili aleatorie i.i.d. con* $\mathbb{E}[Y_i] = \mu$ *e* $0 < \mathrm{Var}(Y_i) = \sigma^2 < \infty$*. Allora, la variabile aleatoria normalizzata*

$$\frac{\sqrt{n}\,(\overline{Y}_n - \mu)}{\sigma}$$

converge in distribuzione a una variabile aleatoria Y *distribuita secondo una normale standard, cioè*

$$\frac{\sqrt{n}\,(\overline{Y}_n - \mu)}{\sigma} \xrightarrow[n\to\infty]{d} Y \sim \mathcal{N}(0,1).$$

Non è difficile verificare che ogni trasformazione affine di un vettore gaussiano $\mathbf{Z}$, ossia una trasformazione del tipo $\mathbf{A}\cdot\mathbf{Z} + \mathbf{b}$, con $\mathbf{A}$ matrice e $\mathbf{b}$ vettore, entrambi costanti, segue anch'essa una distribuzione gaussiana [13]. Quindi, in termini pratici, il teorema afferma che qualsiasi trasformazione affine della somma di un numero elevato di variabili indipendenti tende ad avere una distribuzione approssimativamente normale per n abbastanza grande. Quanto debba essere grande la numerosità n del campione dipende dalla distribuzione da cui vengono campionati i dati. Ad esempio, se la distribuzione della popolazione è normale, allora $\overline{Y}_n$ sarà a sua volta normale indipendentemente da n. Una buona regola empirica è che si può essere confidenti nella validità dell'approssimazione se n è almeno 30 [9, 43]. Le statistiche X e S dell' esempio 2.1 hanno, dunque, una distribuzione asintoticamente normale se gli individui monitorati sono abbastanza numerosi. In realtà, la statistica vettoriale (X, S) ha una distribuzione normale bivariata. Infatti, in generale, se le variabili aleatorie che costituiscono il campione casuale sono a valori vettoriali, il teorema limite centrale continua a valere con opportune modifiche.

Teorema 2.3 (Teorema limite centrale multivariato)**.** *Sia* $\{\mathbf{Y}_1, \mathbf{Y}_2, \ldots\}$ *una successione di variabili aleatorie i.i.d. in* $\mathbb{R}^k$*, con*

$$\mathbb{E}[\mathbf{Y}_i] = \boldsymbol{\mu} \in \mathbb{R}^k, \quad \text{VarCov}(\mathbf{Y}_i) = \boldsymbol{\Sigma} \in \mathbb{R}^{k \times k},$$

dove $\boldsymbol{\Sigma}$ *è una matrice di varianza e covarianza definita positiva e finita.*

Sia

$$\overline{\mathbf{Y}}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{Y}_i$$

la media campionaria vettoriale. Allora, la variabile aleatoria normalizzata

$$\sqrt{n}\,(\overline{\mathbf{Y}}_n - \boldsymbol{\mu})$$

converge in distribuzione a una variabile aleatoria $\mathbf{Y}$ *con distribuzione normale multivariata di media nulla e matrice di varianza e covarianza* $\boldsymbol{\Sigma}$*, cioè*

$$\sqrt{n}\,(\overline{\mathbf{Y}}_n - \boldsymbol{\mu}) \xrightarrow[n\to\infty]{d} \mathbf{Y} \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma}).$$

Il vettore delle medie $\boldsymbol{\mu}$ e la matrice di varianza e covarianza $\boldsymbol{\Sigma}$ della statistica (X, S) verranno calcolate esplicitamente nel prossimo capitolo.

2.2.2 Metodo Delta

La sezione precedente fornisce le condizioni in base alle quali una variabile aleatoria standardizzata ha una distribuzione limite normale. Ci sono tuttavia molti casi in cui non si è interessati specificamente alla distribuzione della variabile aleatoria in sé, ma piuttosto a qualche funzione della variabile aleatoria. Un metodo per procedere consiste nell'utilizzare un'approssimazione tramite serie di Taylor, che consente di approssimare la media e la varianza di una funzione di una variabile aleatoria. Utilizzando queste approssimazioni tramite serie di Taylor per la media e la varianza, si ottiene la seguente utile generalizzazione del Teorema limite centrale, nota come *Metodo Delta*.

Teorema 2.4 (Metodo Delta)**.** *Sia* $\{Y_1, Y_2, \ldots\}$ *una successione di variabili aleatorie tale che*

$$\sqrt{n}\,(Y_n - \theta) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \sigma^2).$$

Per una data funzione g *e un valore specifico di* θ*, si supponga che la derivata* $g'(\theta)$ *esista e sia diversa da zero. Allora*

$$\sqrt{n}\,(g(Y_n) - g(\theta)) \xrightarrow[n\to\infty]{d} \mathcal{N}\big(0,\, \sigma^2[g'(\theta)]^2\big).$$

Analogamente al teorema limite centrale, il metodo Delta si estende naturalmente anche al caso multivariato.

2.3 Stimatori puntuali

Quando il campionamento avviene da una popolazione descritta da una funzione di densità $f(y \mid \theta)$, la conoscenza del parametro θ implica la conoscenza dell'intera popolazione. È quindi naturale cercare un metodo per individuare un buono *stimatore puntuale* di θ, ossia una statistica del campione, la cui realizzazione è detta *stima puntuale* di θ.

2.3.1 Stimatori di massima verosimiglianza

Uno degli approcci più comuni per derivare stimatori puntuali è il metodo di *massima verosimiglianza*.
La funzione di verosimiglianza è uno strumento che consente di valutare quanto un certo insieme di parametri di un modello statistico renda plausibili i dati effettivamente osservati. In altre parole, mentre la funzione di probabilità (o densità) descrive la probabilità di osservare certi dati se i parametri sono noti, la verosimiglianza rovescia il punto di vista e, dato il campione osservato, misura quanto sia "verosimile" ciascun possibile valore dei parametri.

Definizione 2.5. *Sia $f(\mathbf{y} \mid \theta)$ la densità congiunta del campione $\mathbf{Y} = (Y_1, \ldots, Y_n)$. Allora, dato che si è osservato $\mathbf{Y} = \mathbf{y}$, la funzione del parametro θ definita da*

$$\mathcal{L}(\theta \mid \mathbf{y}) = f(\mathbf{y} \mid \theta)$$

si chiama ***funzione di verosimiglianza****.*

L'unica distinzione tra queste due funzioni riguarda quale variabile si considera fissa e quale invece variabile. Quando si considera la funzione di densità $f(\mathbf{y} \mid \theta)$, θ è fisso e $\mathbf{y}$ è variabile; quando invece si considera la funzione di verosimiglianza $\mathcal{L}(\theta \mid \mathbf{y})$, $\mathbf{y}$ rappresenta il campione osservato e θ può variare su tutti i possibili valori dello spazio parametrico.
Il principio di massima verosimiglianza stabilisce che la stima dei parametri debba essere scelta come il valore che rende il campione osservato il più verosimile possibile, ossia quello che massimizza la funzione di verosimiglianza. Per esempio, se si osserva l'esito

di numerosi lanci di una moneta, la funzione di verosimiglianza associa a ciascun valore della probabilità di ottenere testa la "plausibilità" che sia stato proprio quel valore a generare i risultati ottenuti. L'approccio di massima verosimiglianza selezionerà come stima di tale probabilità il valore che rende i dati osservati più coerenti con il modello (la frequenza relativa delle teste in questo caso specifico).

Definizione 2.6. *Per ogni punto campionario* $\mathbf{y}$*, sia* $\hat{\theta}(\mathbf{y})$ *il valore del parametro in cui la funzione di verosimiglianza* $\mathcal{L}(\theta \mid \mathbf{y})$ *raggiunge il suo massimo come funzione di* θ*, con* $\mathbf{y}$ *fissato. Uno* ***stimatore di massima verosimiglianza*** (MLE) *del parametro* θ*, basato su un campione* $\mathbf{Y}$*, è* $\hat{\theta}(\mathbf{Y})$*.*

Quando $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ è multidimensionale, il problema di trovare una stima MLE consiste nel massimizzare una funzione di più variabili. Se la funzione di verosimiglianza è differenziabile rispetto ai parametri incogniti, i possibili candidati per l'MLE sono i valori di $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ che soddisfano

$$\frac{\partial}{\partial \theta_i} \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y}) = 0, \quad i = 1, \ldots, k.$$

I punti in cui le derivate prime si annullano possono corrispondere a minimi locali o globali, massimi locali o globali, oppure a punti di flesso. L'obiettivo è individuare un massimo globale.

Nella maggior parte dei casi, soprattutto quando si può ricorrere alla derivazione, è più semplice lavorare con il logaritmo naturale di $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})$, cioè $\log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})$ (noto come *log-verosimiglianza*), piuttosto che con $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})$ direttamente. Ciò è possibile perché la funzione logaritmo è strettamente crescente su $(0, \infty)$, il che implica, in particolare, che i punti di massimo di $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})$ e di $\log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})$ coincidono.

Esempio 2.2. *Nell'esempio 2.1 la derivata prima della log-verosimiglianza* $\log \mathcal{L}(\lambda \mid \mathbf{y})$ *è data da*

$$\frac{\partial}{\partial \lambda} \log L(\lambda \mid y) = \frac{x}{\lambda} - s,$$

che si annulla nel punto

$$\hat{\lambda}(\mathbf{y}) = \frac{x}{s}.$$

Poichè la derivata seconda della log-versoimiglianza è negativa per ogni valore di $\lambda > 0$*,* $\hat{\lambda}(y)$ *è l'unico punto di massimo globale della funzione e, quindi,*

$$\hat{\lambda}(Y) = \frac{X}{S}$$

è lo stimatore di massima verosimiglianza del parametro esponenziale λ*.*

La procedura di massimazione analizzata si può applicare anche nel caso in cui si è interessati a stimare una trasformazione del parametro o dei parametri incogniti. Infatti, la stima di massima verosimiglianza gode della seguente proprietà di invarianza.

Teorema 2.5 (Proprietà di invarianza)**.** *Se l'*MLE *di* $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ *è* $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$*, e se* $g(\theta_1, \dots, \theta_k)$ *è una qualunque funzione dei parametri, allora l'*MLE *di* $g(\theta_1, \dots, \theta_k)$ *è* $g(\hat{\theta}_1, \dots, \hat{\theta}_k)$*.*

Lo stimatore di massima verosimiglianza di un parametro θ gode di un'altra proprietà notevole. Infatti, sotto opportune condizioni di regolarità (in genere soddisfatte dalle famiglie di distribuzioni esponenziali canoniche [4]), lo stimatore di massima verosimiglianza ha una distribuzione asintoticamente normale.

Teorema 2.6 (Teorema di normalità asintotica dell'MLE)**.** *Siano* $Y_1, \dots, Y_n$ *variabili aleatorie i.i.d. provenienti da una distribuzione avente densità* $f(y \mid \theta)$*. Sia* $\hat{\theta}$ *la stima di massima verosimiglianza di* θ*. Sotto opportune condizioni di regolarità* [4] *per* $f(y \mid \theta)$ *e, quindi, per* $\mathcal{L}(\theta \mid y_1, \dots, y_n)$*, vale che*

$$\sqrt{n}\,(\hat{\theta} - \theta) \xrightarrow[n\to\infty]{d} \mathcal{N}(0,\, v(\theta)^{-1}),$$

dove la quantità

$$v(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ \log f(Y \mid \theta)\right)^2\right]$$

è detta ***informazione di Fisher****.*

Applicando il teorema 2.6 all'esempio 2.2, si ha

$$\sqrt{n}\left(\frac{X}{S} - \lambda\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0,\, v(\lambda)^{-1}),$$

dove

$$v(\lambda) = \frac{\mathbb{E}\left[\mathbb{1}_{\{T<C\}}\right]}{\lambda^2} = \frac{\mathbb{P}(T < C)}{\lambda^2}.$$

Infine, si osservi che il metodo Delta garantisce che qualsiasi trasformazione regolare di uno stimatore di massima verosimiglianza segua una distribuzione asintoticamente normale. Infatti, ad esempio, la trasformzione logaritmica dello stimatore X/S ha la seguente distribuzione asintotica

$$\sqrt{n}\left(\log\left(\frac{X}{S}\right) - \log(\lambda)\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0,\, \frac{1}{\mathbb{P}(T < C)}\right). \tag{2.2}$$

2.3.2 Valutazione degli stimatori puntuali

Esistono diverse misure che permettono di valutare la qualità di uno stimatore di massima verosimiglianza, come quella di un qualsiasi stimatore puntuale ottenuto adottando altre procedure statistiche.

Definizione 2.7. *L'**errore quadratico medio*** (MSE) *di uno stimatore W di un parametro θ è la funzione di θ definita da*

$$\mathrm{MSE}_\theta(W) = \mathbb{E}\Big[(W-\theta)^2\Big].$$

Si noti che l'MSE misura la differenza quadratica media tra lo stimatore W e il parametro θ, una misura ragionevole della bontà di uno stimatore puntuale. In generale, qualsiasi funzione crescente della distanza assoluta $|W-\theta|$ potrebbe essere utilizzata per misurare la bontà di uno stimatore (ad esempio, l'*errore assoluto medio*, $\mathbb{E}_\theta\big[|W-\theta|\big]$, costituisce un'alternativa ragionevole). Tuttavia, l'MSE presenta almeno due vantaggi rispetto ad altre misure di distanza: in primo luogo, è analiticamente più trattabile; in secondo luogo, ammette la seguente interpretazione:

$$\mathbb{E}_\theta\big[(W-\theta)^2\big] = \mathrm{Var}_\theta(W) + \big(\mathbb{E}_\theta[W]-\theta\big)^2 = \mathrm{Var}_\theta(W) + \big(\mathrm{Bias}_\theta(W)\big)^2,$$

dove

$$\mathrm{Bias}_\theta(W) = \mathbb{E}_\theta[W]-\theta.$$

Pertanto, l'MSE incorpora due componenti: una che misura la variabilità dello stimatore (*precisione*) e l'altra che misura la sua distorsione (*accuratezza*). Uno stimatore con buone proprietà in termini di MSE presenta una varianza ridotta e un bias contenuto.

2.4 Test di ipotesi

Si supponga di disporre di un campione aleatorio $\mathbf{y}$ proveniente da una distribuzione che ci è nota a meno di un parametro θ incognito. Un classico test di ipotesi non consiste nello stimare direttamente questo parametro, ma piuttosto nell'utilizzare il campione osservato per verificare qualche ipotesi che lo coinvolga. Un'ipotesi statistica è normalmente un'affermazione su uno o più parametri della distribuzione. Si parla di ipotesi perchè non si sa se sia vera o meno: il problema primario è quello di sviluppare una procedura statistica per determinare se i valori del campione osservato e l'ipotesi fatta

siano compatibili o meno. Se θ denota un parametro della popolazione, la forma generale dell'ipotesi nulla e dell'ipotesi alternativa è $H_0 : \theta \in \Theta_0 \quad$ e $\quad H_1 : \theta \in \Theta_0^c$, dove Θ_0 è un sottoinsieme dello spazio dei parametri e Θ_0^c è il suo complemento. Per esempio, se il parametro di interesse è l'effcacia VE di un vaccino, si potrebbe essere interessati nel testare $H_0 : \text{VE} \leq \text{VE}_0 \quad$ contro $\quad H_1 : \text{VE} > \text{VE}_0$, dove VE_0 è un qualche valore dell'efficacia del vaccino considerato clinicamente rilevante.

Definizione 2.8. *Una **procedura di verifica d'ipotesi**, o **test d'ipotesi**, è una regola che specifica:*

*i. l'insieme, detto **regione di accettazione**, dei valori campionari per cui si decide di accettare H_0 come vera;*

*ii. l'insieme, detto **regione di rifiuto**, dei valori campionari per cui H_0 viene rifiutata e H_1 viene accettata come vera.*

La regione di accettazione è chiaramente l'insieme complementare della regione di rifiuto.
Tipicamente, un test di ipotesi è specificato in termini di una *statistica test* $W(Y_1, \ldots, Y_n) = W(\mathbf{Y})$, cioè una funzione del campione. Ad esempio, un test può specificare che H_0 debba essere rifiutata se $\bar{Y}$, la media campionaria, è maggiore di una certa soglia c. In questo caso la statistica test è $W(\mathbf{Y}) = \bar{Y}$ e la regione di rifiuto è

$$\{(y_1, \ldots, y_n) : \bar{y} > c\}.$$

Un test di ipotesi della forma $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_0^c$ può incorrere in uno di due tipi di errore, che sono tradizionalmente chiamati *Errore di tipo I* ed *Errore di tipo II*. Se $\theta \in \Theta_0$, ma la statistica test su cui si basa la regola decisionale porta erroneamente a rifiutare H_0, allora il test commette un errore di tipo I. Viceversa, se $\theta \in \Theta_0^c$, ma la statistica test porta ad accettare (o non rifiutare) H_0, allora il test commette un errore di tipo II. Più precisamente, se R è la regione di rifiuto di un test, per $\theta \in \Theta_0$, il test commette un errore se $\mathbf{y} \in R$, per cui la probabilità di un errore di tipo I è $\mathbb{P}(\mathbf{Y} \in R \mid \theta)$; invece, per $\theta \in \Theta_0^c$, la probabilità di un errore di tipo II è $\mathbb{P}(\mathbf{Y} \in R^c \mid \theta)$. Dunque, la funzione di θ, $\mathbb{P}(\mathbf{Y} \in R \mid \theta)$, contiene tutte le informazioni sul test con regione di rifiuto R, ovvero:

$$\mathbb{P}(\mathbf{Y} \in R \mid \theta) = \begin{cases} \text{probabilità di un errore di tipo I,} & \text{se } \theta \in \Theta_0, \\ 1 - \text{probabilità di un errore di tipo II,} & \text{se } \theta \in \Theta_0^c. \end{cases}$$

Questa considerazione porta alla seguente definizione.

Definizione 2.9. *La **potenza** di un test d'ipotesi con regione di rifiuto R è la funzione di θ definita da*

$$\beta(\theta) = \mathbb{P}(\mathbf{Y} \in R \mid \theta).$$

La funzione potenza ideale è pari a zero per tutti i valori di $\theta \in \Theta_0$ e pari a uno per tutti i valori di $\theta \in \Theta_0^c$. Tale potenza ideale non può essere raggiunta, eccetto in casi degeneri. Qualitativamente, un buon test ha una funzione di potenza vicina a uno per la maggior parte dei valori di $\theta \in \Theta_0^c$ e vicina a zero per la maggior parte dei valori di $\theta \in \Theta_0$.
Tipicamente, la potenza di un test dipende dalla dimensione campionaria n. Se n è un parametro di controllo, lo studio della funzione potenza può aiutare a determinare quale dimensione campionaria sia appropriata in un esperimento. Per una dimensione campionaria fissata, è di solito impossibile rendere entrambe le probabilità di errore arbitrariamente piccole. Nella ricerca di un buon test, è comune limitarsi a considerare quei test che controllano la probabilità di errore di tipo I a un livello specificato. All'interno di questa classe di test si cercano poi quelli che hanno la probabilità di errore di tipo II più piccola possibile. La seguente definizione risulta utile quando si discutono test che controllano le probabilità di errore di tipo I.

Definizione 2.10. *Per $0 \leq \alpha \leq 1$, un test avente potenza $\beta(\theta)$ è detto **test con livello di significatività α** se*

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha.$$

Comunemente si specifica il livello del test che si intende utilizzare, con scelte tipiche che sono $\alpha = 0.01,\ 0.025,\ 0.05,$ e 0.10. Occorre notare che, fissando il livello del test, chi lo conduce sta controllando solo le probabilità di errore di tipo I, non quelle di tipo II. Se si adotta questo approccio, si dovrebbero specificare le due ipotesi, nulla e alternativa, in modo tale che sia più importante controllare la probabilità di errore di tipo I. Si consideri, ad esempio, il caso di un ricercatore che si aspetti che un esperimento confermi una determinata ipotesi, ma che desideri dichiararla valida solo se i dati forniscono un sostegno realmente solido. In questo contesto, il test statistico può essere impostato in modo che l'ipotesi alternativa coincida con l'ipotesi che ci si attende venga supportata dai dati e che si intende dimostrare. Proprio per questo, l'ipotesi alternativa viene talvolta definita *ipotesi di ricerca* (ad esempio, l'efficacia di un vaccino). L'adozione di un test

con livello di significatività α sufficientemente piccolo consente al ricercatore di ridurre il rischio di affermare erroneamente che i dati sostengano l'ipotesi di ricerca quando, in realtà, essa non è vera.

2.5 Intervalli di confidenza

Nella stima puntuale di un parametro θ, l'inferenza consiste nell'individuare un singolo valore come stima di θ. Invece, in un problema in cui si intende stimare un insieme di valori plausibili del parametro θ, l'inferenza consiste nell'affermare che

$$\theta \in I,$$

dove $I \subset \Theta$ e $I = I(\mathbf{y})$ è un insieme determinato dal valore del campione osservato $\mathbf{Y} = \mathbf{y}$. Se θ è un parametro reale, si preferisce di solito che l'insieme di stima I sia un intervallo.

Definizione 2.11. *Una* ***stima intervallare*** *di un parametro reale θ è costituita da una coppia di funzioni $L(y_1, \ldots, y_n)$ e $U(y_1, \ldots, y_n)$, dipendenti dal campione, tali che $L(\mathbf{y}) \leq U(\mathbf{y})$ per ogni valore campionario $\mathbf{y}$. Se si osserva $\mathbf{Y} = \mathbf{y}$, si formula l'inferenza $L(\mathbf{y}) \leq \theta \leq U(\mathbf{y})$. L'intervallo aleatorio $[L(\mathbf{Y}), U(\mathbf{Y})]$ è detto* ***stimatore intervallare****.*

Lo scopo di utilizzare uno stimatore intervallare, piuttosto che uno stimatore puntuale, è quello di avere una certa garanzia di includere il parametro di interesse. Il grado di certezza di tale garanzia è quantificato nelle seguenti definizioni.

Definizione 2.12. *Per uno stimatore intervallare $[L(\mathbf{Y}), U(\mathbf{Y})]$ di un parametro θ, la* ***probabilità di copertura*** *di $[L(\mathbf{Y}), U(\mathbf{Y})]$ è la probabilità che l'intervallo casuale $[L(\mathbf{Y}), U(\mathbf{Y})]$ contenga il vero valore del parametro θ. In simboli, essa si denota con*

$$\mathbb{P}(\theta \in [L(\mathbf{Y}), U(\mathbf{Y})] \mid \theta)\,.$$

Definizione 2.13. *Per uno stimatore intervallare $[L(\mathbf{Y}), U(\mathbf{Y})]$ di un parametro θ, il* ***coefficiente di confidenza*** *di $[L(\mathbf{Y}), U(\mathbf{Y})]$ è l'estremo inferiore delle probabilità di copertura:*

$$\inf_{\theta} \mathbb{P}(\theta \in [L(\mathbf{Y}), U(\mathbf{Y})] \mid \theta)\,.$$

Uno stimatore intervallare con un coefficiente di confidenza $1 - \alpha$ è detto ***intervallo di confidenza*** *$1 - \alpha$.*

Più in generale, quando non si è certi della forma esatta dell'insieme con cui si stima il parametro di interesse, si parla di *insieme di confidenza*.
È importante tenere presente che la quantità aleatoria è l'intervallo (insieme), non il parametro. Pertanto, quando si scrivono enunciati di probabilità come $\mathbb{P}\big(\theta \in [L(\mathbf{Y}), U(\mathbf{Y})] \mid \theta\big)$, tali enunciati si riferiscono al campione $\mathbf{Y}$, non al parametro θ. Quindi, l'espressione $\mathbb{P}\big(\theta \in [L(\mathbf{Y}), U(\mathbf{Y})] \mid \theta\big)$, che potrebbe sembrare un'affermazione su un θ casuale, va intesa come l'espressione algebricamente equivalente

$$\mathbb{P}(L(\mathbf{Y}) \leq \theta,\ U(\mathbf{Y}) \geq \theta \mid \theta),$$

cioè un'affermazione sul vettore aleatorio $\mathbf{Y}$.
Esiste una corrispondenza molto stretta tra la verifica d'ipotesi e la stima intervallare. Infatti, in generale, ad ogni insieme di confidenza corrisponde un test e, viceversa, ad ogni test corrisponde un insieme di confidenza. Infatti, poiché entrambe le procedure cercano coerenza tra le statistiche campionarie e i parametri della popolazione, i test e gli insiemi di confidenza pongono la stessa domanda, ma da una prospettiva leggermente diversa. Il test d'ipotesi fissa il valore del parametro e si chiede quali valori campionari (la regione di accettazione) siano coerenti con quel valore fissato. L'insieme di confidenza, invece, fissa il valore campionario e si chiede quali valori del parametro (l'intervallo di confidenza) rendano tale valore campionario il più plausibile possibile. La corrispondenza tra le regioni di accettazione dei test e gli insiemi di confidenza è garantita in generale dal seguente teorema.

Teorema 2.7. *Per ogni valore parametrico $\theta_0 \in \Theta$, sia $A(\theta_0)$ la regione di accettazione di un test di livello α per $H_0 : \theta = \theta_0$. Per ogni valore campionario $\mathbf{y} \in \mathcal{Y}$, sia $C(\mathbf{y})$ un sottoinsieme dello spazio dei parametri della forma*

$$I(\mathbf{y}) = \{\theta_0 : \mathbf{y} \in A(\theta_0)\}.$$

Allora l'insieme aleatorio $I(\mathbf{Y})$ è un insieme di confidenza di livello $1 - \alpha$. Viceversa, sia $I(\mathbf{Y})$ un insieme di confidenza di livello $1 - \alpha$. Per ogni $\theta_0 \in \Theta$, si definisca

$$A(\theta_0) = \{\mathbf{y} : \theta_0 \in I(\mathbf{y})\}.$$

Allora $A(\theta_0)$ è la regione di accettazione di un test di livello α per $H_0 : \theta = \theta_0$.

Sebbene sia comune parlare di inversione di un test per ottenere un insieme di confidenza, il Teorema 2.7 chiarisce che in realtà si dispone di una famiglia di test, uno per

ciascun valore di $\theta_0 \in \Theta$, che viene invertita per ottenere un singolo insieme di confidenza. Inoltre, nel teorema si enuncia soltanto l'ipotesi nulla $H_0 : \theta = \theta_0$. Tutto ciò che è richiesto alla regione di accettazione è che soddisfi la disuguaglianza

$$\mathbb{P}\big(\mathbf{Y} \in A(\theta_0) \mid \theta_0\big) \;\geq\; 1 - \alpha.$$

Tuttavia, in pratica, quando si costruisce un insieme di confidenza tramite inversione del test, si ha anche in mente un'ipotesi alternativa, come ad esempio un'ipotesi *bilaterale* $H_1 : \theta \neq \theta_0$ oppure un'ipotesi *unilaterale* $H_1 : \theta > \theta_0$. L'alternativa scelta determinerà la forma ragionevole di $A(\theta_0)$ e la forma di $A(\theta_0)$ determinerà a sua volta la forma di $I(\mathbf{y})$. Si noti, tuttavia, che non vi è alcuna garanzia che l'insieme di confidenza ottenuto tramite inversione del test sia effettivamente un intervallo. Nella maggior parte dei casi, tuttavia, i test unilaterali generano intervalli unilaterali, i test bilaterali generano intervalli bilaterali e regioni di accettazione dalla forma irregolare producono insiemi di confidenza altrettanto irregolari.

2.5.1 Metodo di Clopper-Pearson

Il metodo Clopper-Pearson è una tecnica di *inferenza pivotale* che consente di costruire un intervallo di confidenza per un parametro θ, con livello di copertura almeno $1-\alpha$, sulla base di una variabile aleatoria $T(\mathbf{Y}|\theta)$ dipendente dal campione di dati $\mathbf{Y} = (Y_1, \ldots, Y_n)$ e dal parametro θ.

Definizione 2.14. *Una variabile aleatoria $T(\mathbf{Y}, \theta) = T(Y_1, \ldots, Y_n, \theta)$ si dice **quantità pivotale** (o impropriamente **statistica pivotale**) se la distribuzione di $T(Y, \theta)$ è indipendente da θ. In altre parole, se $\mathbf{Y} \sim f(y_1, \ldots, y_n \mid \theta)$, allora $T(\mathbf{Y}, \theta)$ ha la stessa distribuzione per tutti i valori di θ.*

Se $F_T(t|\theta)$ è la funzione di ripartizione di $T(\mathbf{Y}, \theta)$, l'intervallo di confidenza $1 - \alpha$ si ottiene invertendo la regione di accettazione di un test statistico con livello di significatività α la cui regola di rifiuto (e quindi di accettazione) coinvolge F_T. Tale regola è definita in modo tale che la regione di accettazione sia un insieme aleatorio che copra il valore vero del parametro θ con probabilità $1 - \alpha$. Invertendo questa regione, si ottiene un insieme di valori plausibili per θ con livello di confidenza $1 - \alpha$, che risulta essere un intervallo se la funzione di ripartizione è monotona in θ per ogni valore assumibile dalla statistica T [4].

In particolare, se T è una variabile aleatoria continua, si verifica che la variabile aleatoria $F_T(T|\theta)$ è distribuita uniformemente sull'intervallo (0,1) ed è, quindi, una statistica pivotale che può essere utilizzata per costruire un insieme aleatorio con probabilità di copertura esattamente $1-\alpha$ [4]. Al contrario, se T è discreta, la variabile aleatoria $F_T(T|\theta)$ non è necessariamente una quantità pivotale. Nonostante ciò, si può comunque costruire un insieme aleatorio con probabilità di copertura almeno $1-\alpha$, che risulta in generale più conservativo. In questo caso, la regione di accettazione viene definita combinando la funzione di ripartizione della variabile T con quella della variabile $-T$ e sfruttando il fatto che, data una generica variabile aleatoria discreta Z con funzione di ripartizione $F_Z(z)$, la variabile aleatoria $F_Z(Z)$ domina stocasticamente una variabile U uniformemente distribuita sull'intervallo (0,1) [4]. Formalmente, $\forall u \in (0,1)$, la dominanza stocastica è espressa dalla disuguaglianza

$$\mathbb{P}(F_Z(Z) <= u) <= \mathbb{P}(U <= u) = u,$$

che è equivalente alla disuguaglianza

$$\mathbb{P}(F_Z(Z) > u) >= \mathbb{P}(U > u) = 1 - u.$$

Teorema 2.8 (Pivoting di una funzione di ripartizione discreta)**.** *Sia T una statistica discreta con funzione di ripartizione $F_T(t|\theta) = \mathbb{P}(T <= t|\theta)$ decrescente in θ per ogni realizzazione t della statistica. Siano $\alpha_1, \alpha_2 > 0$ due valori fissati tali che $\alpha_1 + \alpha_2 = \alpha \in (0,1)$ e si definiscano, per ogni t, i valori $\theta_L(t), \theta_U(t)$ tali che*

$$\mathbb{P}(T <= t|\theta_U(t)) = \alpha_1, \quad \mathbb{P}(T >= t|\theta_L(t)) = \alpha_2. \tag{2.3}$$

Allora, l'intervallo aleatorio $[\theta_L(T), \theta_U(T)]$ è un intervallo di confidenza con livello di copertura almeno $1-\alpha$.

Osservazione 2.1. *Se $F_T(t|\theta)$ è crescente in θ per ogni t, con opportune modifiche nella formula (2.3), il* Teorema 2.8 *continua a valere.*

2.5.1.1 Intervallo di confidenza bilaterale nel caso binomiale

Si consideri un campione aleatorio $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Bern}(\theta)$ e la relativa statistica $T = \sum_{i=1}^{n} Y_i \sim \text{Binom}(n, \theta)$. Si supponga di voler utilizzare il metodo Clopper-Pearson per costruire un intervallo bilaterale di confidenza $1-\alpha$ per il parametro θ ripartendo α

equamente sulle code, cioè ponendo $\alpha_1 = \alpha_2 = \alpha/2$. Sebbene questa non sia sempre la scelta ottimale, è sicuramente una strategia ragionevole in molte situazioni. Per ottenere tale intervallo, è necessario il seguente risultato preliminare che lega la funzione di ripartizione binomiale con la funzione di ripartizione di una distribuzione beta.

Proposizione 2.1. *Sia $B(t)$ una variabile aleatoria distribuita secondo una Beta con parametri t e $n-t+1$, ovvero*

$$B(t) \sim \text{Beta}(t, n-t+1).$$

Allora, per ogni $\theta \in [0,1]$, vale la seguente uguaglianza:

$$\mathbb{P}(T \geq t \mid \theta) = \mathbb{P}(B(t) \leq \theta). \tag{2.4}$$

Osservando che la funzione di ripartizione binomiale è decrescente in θ per ogni valore della statistica T e applicando la Proposizione 2.1 alla definizione (2.3) che compare nell'enunciato del Teorema 2.8, per l'estremo inferiore dell'intervallo si ottiene

$$F_{B(t)}(\theta_L(t)) = \alpha/2 \quad \Rightarrow \quad \theta_L(t)) = F^{-1}_{B(t)}(\alpha/2) = \text{Beta}_{\alpha/2}(t, n-t+1),$$

mentre per l'estremo superiore si ha

$$\mathbb{P}(T < t+1 | \theta_U(t)) = \alpha/2 \quad \Rightarrow \quad F_{B(t+1)}(\theta_U(t)) = 1-\alpha/2$$

$$\Rightarrow \theta_U(t)) = F^{-1}_{B(t+1)}(\alpha/2) = \text{Beta}_{1-\alpha/2}(t+1, n-t).$$

L'intervallo $[\theta_L(t),\ \theta_u(t)]$ è noto in letteratura come *intervalllo di confidenza esatto di Clopper-Pearson* (CP), in quanto il metodo di Clopper-Pearson, applicato a una statistica binomiale, permette di ottenere una probabilità di copertura che non si basa su approssimazioni asintotiche, ma coincide (o supera) esattamente il livello nominale.

2.5.2 Intervalli di confidenza asintotici per rapporti

Si supponga di essere interessati a stimare il rapporto

$$R = \frac{\theta}{\tau}$$

di due parametri θ e τ, i cui stimatori di massima verosimiglianza sono rispettivamente $\hat{\theta}$ e $\hat{\tau}$. Sebbene i due stimatori abbiano entrambi distribuzioni asintoticamente normali

in virtù del teorema 2.6, lo stesso non si può dire per il loro rapporto. Tuttavia, per campioni sufficientemente grandi, il teorema 2.4 applicato ad entrambi gli stimatori garantisce che

$$\log\left(\frac{\hat{\theta}}{\hat{\tau}}\right) = \log(\hat{\theta}) - \log(\hat{\tau}) \sim \mathcal{N}\left(\log(R),\ \frac{v_\theta(\theta)^{-1}}{n\theta^2} + \frac{v_\tau(\tau)^{-1}}{m\tau^2}\right), \tag{2.5}$$

dove n è la dimensione del campione estratto dalla distribuzione parametrica dipendente da θ, avente informazione di Fisher $v_\theta(\cdot)$, e m è la dimensione del campione estratto dalla distribuzione parametrica dipendente da τ, avente informazione di Fisher $v_\tau(\cdot)$. Disponendo di una stima $\hat{\sigma}^2$ della varianza della distribuzione normale (2.5), si può costruire un intervallo di confidenza asintotico per il parametro $\log(R)$ con copertura $1-\alpha$, cioè

$$\log\left(\frac{\hat{\theta}}{\hat{\tau}}\right) - \hat{\sigma} z_{1-\alpha/2} \leq \log(R) \leq \log\left(\frac{\hat{\theta}}{\hat{\tau}}\right) + \hat{\sigma} z_{1-\alpha/2}, \tag{2.6}$$

dove $z_{1-\alpha/2}$ è il quantile $1-\alpha/2$ della normale standard. Applicando la trasformazione esponenziale ad entrambi gli estremi dell'intervallo (2.6), si ottiene un intervallo di confidenza con copertura (approssimativamente) $1-\alpha$ per il rapporto R.

2.6 Statistica bayesiana

Nell'approccio statistico analizzato sinora, il cosiddetto *approccio frequentista*, il parametro θ è considerato come una quantità incognita ma fissa. Un campione casuale $Y_1, \ldots, Y_n$ è estratto da una popolazione indicizzata da θ e, sulla base dei valori osservati nel campione, si ottengono informazioni sul valore di θ.

Nell'*approccio bayesiano*, invece, θ è considerato come una quantità la cui variabilità può essere descritta da una distribuzione di probabilità, detta *distribuzione a priori* (*prior*). Questa è una distribuzione soggettiva, basata sulle convinzioni di chi conduce l'esperimento, e viene formulata prima di osservare i dati (da qui il nome "a priori"). Successivamente, si osserva un campione da una popolazione indicizzata da θ e la distribuzione a priori viene aggiornata con l'informazione fornita dal campione. La distribuzione aggiornata prende il nome di *distribuzione a posteriori* (*posterior*). Tuttavia, in alcune varianti, come nell'approccio *Empirical Bayes*, la prior non viene fissata prima dell'osservazione ma stimata a partire dai dati stessi, rendendo questo metodo un compromesso tra l'impostazione bayesiana classica e quella frequentista. In ogni caso, l'aggiornamento della prior si effettua tramite la *regola di Bayes*, da cui deriva la denominazione di

statistica bayesiana. Nello specifico, se si denota la distribuzione a priori con $\pi(\theta)$ e la distribuzione campionaria con $f(\mathbf{y} \mid \theta)$ (ovvero la funzione di verosimiglianza), allora la distribuzione a posteriori, cioè la distribuzione condizionata di θ dato il campione osservato $\mathbf{y}$, è

$$\pi(\theta \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \theta)\,\pi(\theta)}{m(\mathbf{y})},$$

dove $m(\mathbf{y})$ è la distribuzione marginale di $\mathbf{Y}$, ossia

$$m(\mathbf{y}) = \int f(\mathbf{y} \mid \theta)\,\pi(\theta)\,d\theta.$$

Si noti che, nel contesto bayesiano, si usa una lettera minuscola sia per denotare il parametro aleatorio sia la sua realizzazione.

2.6.1 Stimatori bayesiani

La distribuzione a posteriori viene utilizzata per formulare affermazioni sulla quantità aleatoria θ. Ad esempio, la media della distribuzione a posteriori può essere utilizzata come stima puntuale di θ. Infatti, quando si conosce la distribuzione di una variabile aleatoria W, la migliore stima del suo valore, in termini di errore quadratico medio, è fornita dalla media, cioè

$$\mathbb{E}\left[(W - \mathbb{E}[W])^2\right] \leq \mathbb{E}[(W - a)^2], \quad \forall a \in \mathbb{R},$$

a patto che $\mathbb{E}[W]$ sia finito [4]. Dunque, la migliore stima di θ, osservato il campione $\mathbf{Y} = \mathbf{y}$, è data dalla media a posteriori.

$$\mathbb{E}[\theta \mid \mathbf{Y} = \mathbf{y}],$$

che per questo motivo rientra nella categoria degli *stimatori bayesiani*.
Invece, se l'errore di previsione è misurato in termini di errore assoluto medio, la miglior stima di θ, osservati i dati, è fornita dalla mediana a posteriori, che è quindi un altro esempio di stimatore bayesiano [4].

2.6.2 Intervalli di credibilità

Quando si parla di intervalli di confidenza, per sottolineare che la quantità aleatoria è l'intervallo e non il parametro, si afferma che lo stimatore intervallare copre il parametro e non che il parametro cade dentro l'intervallo con una certa probabilità.

Al contrario, l'impostazione bayesiana consente di affermare che θ appartiene a un dato intervallo con una certa probabilità. Ciò è possibile perché, nell'approccio bayesiano, θ è una variabile aleatoria dotata di una distribuzione di probabilità. Tutte le affermazioni bayesiane relative alla copertura sono formulate con riferimento alla distribuzione a posteriori del parametro.

Poiché gli intervalli bayesiani e gli intervalli frequentisti esprimono valutazioni di probabilità di natura molto diversa, per mantenere chiara la distinzione tra i due tipi di insiemi, gli stimatori intervallari bayesiani vengono denominati *intervalli di credibilità*, anziché intervalli di confidenza. In particolare, se $\pi(\theta \mid \mathbf{y})$ è la distribuzione a posteriori di θ dato $\mathbf{Y} = \mathbf{y}$, allora, per ogni intervallo $A \subset \Theta$, la *probabilità credibile* di A è definita come

$$\mathbb{P}(\theta \in A \mid \mathbf{y}) \;=\; \int_A \pi(\theta \mid \mathbf{y})\, d\theta. \tag{9.2.18}$$

È importante non confondere la *probabilità credibile* (la probabilità a posteriori bayesiana) con la *probabilità di copertura* (la probabilità frequentista). Si tratta infatti di concetti molto diversi, con significati e interpretazioni differenti. La probabilità credibile deriva dalla distribuzione a posteriori, che a sua volta ottiene la propria probabilità dalla distribuzione a priori, oltre che dai dati. Pertanto, la probabilità credibile riflette le convinzioni soggettive dello sperimentatore, come espresse nella distribuzione a priori e aggiornate con i dati nella distribuzione a posteriori. Un'affermazione bayesiana di copertura al 90% significa che lo sperimentatore, dopo aver combinato la conoscenza a priori con i dati osservati, è sicuro al 90% che l'intervallo di credibilità contenga il parametro.

La probabilità di copertura, invece, riflette l'incertezza dovuta alla procedura di campionamento, ottenendo la sua probabilità dal meccanismo oggettivo delle ripetizioni sperimentali. Un'affermazione frequentista di copertura al 90% significa che, in una lunga sequenza di prove sperimentali identiche, il 90% degli intervalli di confidenza realizzati conterrà il vero valore del parametro.

Poichè la scelta della distribuzione a priori introduce un certo grado di soggettività nell'approccio bayesiano, ciò può a sua volta determinare un bias indesiderato nella stima a posteriori del parametro di interesse, compromettendo la giustificazione stessa della procedura bayesiana e rendendo più difficile l'interpretazione delle relative affermazioni probabilistiche. Per assicurarsi che ciò non avvenga, la cosiddetta *calibrazione frequentista* è stata proposta come buona pratica nell'implementazione delle procedure

bayesiane. In questo senso, esse vengono considerate *calibrate* (in senso frequentista) se le loro affermazioni probabilistiche possiedono la copertura dichiarata nelle ripetizioni dell'esperimento. Per maggiori dettagli si rimanda a [19].

2.6.3 Test di ipotesi bayesiani

I test d'ipotesi possono essere formulati anche in ambito bayesiano. Infatti, la distribuzione a posteriori consente di calcolare le probabilità che l'ipotesi nulla H_0 e l'ipotesi alternativa H_1 siano vere. Tali probabilità, denotate rispettivamente con $\mathbb{P}(\theta \in \Theta_0 \mid \mathbf{y}) = \mathbb{P}(H_0 \text{ è vera} \mid \mathbf{y})$ e $\mathbb{P}(\theta \in \Theta_0^c \mid \mathbf{y}) = \mathbb{P}(H_1 \text{ è vera} \mid \mathbf{y})$, non sono significative nella statistica frequentista perchè il parametro θ è incognito ma fisso. Di conseguenza, un'ipotesi è o vera o falsa. Se $\theta \in \Theta_0$, $\mathbb{P}(H_0 \text{ è vera} \mid \mathbf{y}) = 1$ e $\mathbb{P}(H_1 \text{ è vera} \mid \mathbf{y}) = 0$ qualsiasi sia il campione osservato $\mathbf{y}$. Se $\theta \in \Theta_0^c$, i precedenti valori sono invertiti. Poichè queste probabilità non sono note (θ è incognito) e non dipendono dal campione $\mathbf{y}$, non si rivelano utili nella statistica frequentista. Tuttavia, in una formulazione bayesiana dei test di ipotesi, queste probabilità dipendono dal campione $\mathbf{y}$ e forniscono informazioni significative sulla veridicità di H_0 e H_1. Infatti, una regola decisionale bayesiana può basarsi sulla distribuzione a posteriori. Per esempio, si rifiuta H_0 se $\mathbb{P}(\theta \in \Theta_0 \mid \mathbf{y}) < \mathbb{P}(\theta \in \Theta_0^c \mid \mathbf{y})$. Usando la terminologia frequentista, la statistica test (funzione del campione) è $\mathbb{P}(\theta \in \Theta_0^c \mid \mathbf{y})$ e la regione di rifiuto è l'insieme

$$R = \left\{ \mathbf{y} : P(\theta \in \Theta_0^c \mid \mathbf{y}) > \tfrac{1}{2} \right\}.$$

Tuttavia, per controllare l'errore di prima specie, ovvero la probabilità di rifiutare H_0 dato che H_0 è vera, si può decidere di rifiutare l'ipotesi nulla solo se $\mathbb{P}(\theta \in \Theta_0^c \mid \mathbf{y})$ è maggiore di una certa soglia probabilistica, scelta in modo tale che l'errore non sia superiore a un certo valore (di solito entro il 2.5% in un test unilaterale).

2.6.4 Distribuzioni a priori coniugate

In generale, fissata una distribuzione campionaria (funzione di verosimiglianza) e scelta una famiglia di distribuzioni a priori, non sempre si può ottenere un'espressione in forma chiusa per la distribuzione a posteriori. Dal punto di vista computazionale, il caso più semplice è quello in cui la distribuzione a posteriori appartiene alla stessa famiglia della distribuzione a priori.

Definizione 2.15. *Sia $\mathcal{F}$ la classe delle funzioni di densità $f(y \mid \theta)$ indicizzate da θ. Una classe Π di distribuzioni a priori è detta* ***famiglia coniugata*** *per $\mathcal{F}$ se la distribuzione a posteriori appartiene a Π per ogni $f \in \mathcal{F}$, per ogni prior in Π e per ogni $y \in \mathcal{Y}$, dove $\mathcal{Y}$ è lo spazio campionario.*

Esempio 2.3. *Siano $Y_1, \ldots, Y_n \overset{iid}{\sim}$ Bernoulli(p). Allora $T = \sum_{i=1}^{n} Y_i \sim$ Binomiale(n, p). Assumendo che la distribuzione a priori di p sia* Beta(a, b)*, la distribuzione congiunta di T e p si ottiene moltiplicando la densità condizionata $f(t \mid p)$ di T per la densità a priori di p, cioè*

$$f(t,p) = \binom{n}{t} p^t (1-p)^{n-t} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\; p^{a-1}(1-p)^{b-1},$$

$$= \binom{n}{t} \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, p^{t+a-1}(1-p)^{n-t+b-1}.$$

La densità marginale di T è

$$f(t) \;=\; \int_0^1 f(t,p)\,dp \;=\; \binom{n}{t} \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, \frac{\Gamma(t+a)\,\Gamma(n-t+b)}{\Gamma(n+a+b)},$$

una distribuzione nota come beta–binomiale*.*
Di conseguenza, la distribuzione a posteriori, ossia la densità condizionata di p dato $T = t$, è

$$f(p \mid t) = \frac{f(t,p)}{f(t)} = \frac{\Gamma(n+a+b)}{\Gamma(t+a)\,\Gamma(n-t+b)}\; p^{\,t+a-1}\,(1-p)^{\,n-t+b-1},$$

che corrisponde a una distribuzione Beta$(t+a,\, n-t+b)$*.*

Dall'esempio precedente si evince che la famiglia Beta è coniugata per la famiglia binomiale. Di conseguenza, se si parte da una prior Beta, si ottiene come risultato una posterior anch'essa Beta. L'aggiornamento della prior si traduce quindi in una semplice modifica dei suoi parametri. Dal punto di vista matematico, ciò è estremamente conveniente, poiché rende i calcoli generalmente più agevoli. Questa proprietà, come si vedrà nel prossimo capitolo, è alla base di uno dei metodi statistici standard adottati nei trial clinici sui vaccini.

2.6.5 Metodi MCMC

Stabilire se una famiglia coniugata sia o meno appropriata per un determinato problema rimane una decisione soggettiva. Le famiglie coniugate sono quelle che garantiscono sempre una distribuzione a posteriori della stessa forma funzionale della distribuzione a priori. Tuttavia, anche con prior non coniugate può capitare che la posterior si scriva comunque in forma chiusa in termini di distribuzioni note. Non sempre però è così: in molti casi la posterior non ha una forma chiusa semplice e bisogna ricorrere a una classe di metodi approssimativi nota come *Metodi di Monte Carlo via Catene di Markov* (MCMC). Casi particolari di tali metodi sono il *Gibbs Sampler* e l'*Algoritmo di Metropolis-Hastings*. Entrambi permettono di ottenere campioni da una qualsiasi distribuzione multivariata (in particolare la distribuzione a posteriori di un parametro). Nel seguito viene enunciato l'algoritmo di Metropolis-Hastings [31].

Teorema 2.9 (Algoritmo di Metropolis–Hastings)**.** *Sia $Y \sim f_Y(y)$ una variabile aleatoria con densità bersaglio f_Y. Sia $q(\cdot \mid z)$ una famiglia di densità di proposta, ciascuna definita sullo stesso supporto di f_Y.*

Per generare $Y \sim f_Y$:

1. *Inizializza con un valore Z_0.*
2. *Per $i = 1,2,\ldots$:*
   (a) *Genera una proposta $V_i \sim q(\cdot \mid Z_{i-1})$.*
   (b) *Genera $U_i \sim \text{Uniform}(0,1)$.*
   (c) *Calcola*
   $$\rho_i \;=\; \min\left\{1,\; \frac{f_Y(V_i)}{f_Y(Z_{i-1})} \cdot \frac{q(Z_{i-1} \mid V_i)}{q(V_i \mid Z_{i-1})}\right\}.$$
   (d) *Aggiorna*
   $$Z_i \;=\; \begin{cases} V_i, & \text{se } U_i \le \rho_i, \\ Z_{i-1}, & \text{se } U_i > \rho_i. \end{cases}$$

Al crescere di i, la sequenza $\{Z_i\}$ converge in distribuzione a $Y \sim f_Y$.

L'algoritmo non produce immediatamente una variabile aleatoria con distribuzione esattamente pari a f_Y, ma genera piuttosto una successione convergente di variabili aleatorie in cui la distribuzione di ciascuna variabile dipende solo dal valore immediatamente

precedente (nota come *Catena di Markov*). In pratica, dopo che l'algoritmo è stato eseguito per un certo numero di iterazioni (ovvero per valori sufficientemente grandi di i), le variabili Z_i prodotte si comportano in modo molto simile a variabili distribuite secondo f_Y. Inoltre, poichè ogni valore di Z_i dipende dal valore precedente Z_{i-1}, i valori della sequenza generata sono correlati e dipendono dall'inizializzazione.
Per ottenere una sequenza di valori indipendente dal valore iniziale, si adotta la procedura di *burn-in*, che consiste nello scartare i primi valori della sequenza generata dall'algoritmo.
Per ridurre la correlazione, invece, si utilizza la procedura di *thinning*, ovvero si selezionano i valori che si trovano a una certa distanza predefinita nella catena. Infatti, se interpretiamo la catena come una sequenza temporale, la correlazione tra i valori generati dipende dalla loro distanza temporale. All'aumentare della distanza diminuisce la correlazione. Un modo per misurare tale correlazione è calcolare la cosiddetta *funzione di autocorrelazione* (ACF) *al lag* h:

$$r(h) = \frac{\sum_{i=1}^{N-h}(z_i - \bar{z})(z_{i+h} - \bar{z})}{\sum_{i=1}^{N}(z_i - \bar{z})^2},$$

dove $z_1, \ldots, z_N$ è la sequenza generata e $\bar{z}$ è la relativa media campionaria. La funzione $r(h)$ valuta la correlazione tra i valori originali e quelli traslati di un tempo (lag) h. Se, per esempio, l'ACF si annulla per $h \geq c$, allora basta adottare un fattore di thinning pari a c per avere un campione con valori indipendenti.
Sia l'algoritmo di Metropolis che quello di Gibbs sono implementati in `JAGS` (*Just Another Gibbs Sampler* [38]), un software open source largamente utilizzato per l'inferenza bayesiana tramite metodi MCMC.
`JAGS` consente di specificare modelli statistici in un linguaggio dichiarativo, simile a quello di `BUGS` (*Bayesian inference Using Gibbs Sampling* [38]), e di generare campioni dalla distribuzione a posteriori dei parametri di interesse mediante algoritmi di tipo Metropolis–Hastings e Gibbs. Grazie a queste funzionalità, `JAGS` è uno strumento flessibile e diffuso nella pratica statistica, particolarmente utile per problemi complessi in cui le distribuzioni a posteriori non hanno forma chiusa e richiedono procedure computazionali di approssimazione.

Parte II

Capitolo 3

Analisi statistica dell'efficacia vaccinale

3.1 Introduzione alla sperimentazione clinica dei vaccini

L'immunizzazione rappresenta uno dei grandi progressi della sanità pubblica. Il primo successo rilevante, nonché l'origine stessa della parola "vaccinazione" (dal latino *vacca*, cioè mucca), risale alla fine del XVIII secolo, quando Jenner introdusse il vaccino basato sul vaiolo bovino contro il vaiolo umano.
Dopo quasi un secolo di pausa, alla fine del XIX secolo furono sviluppati i vaccini contro colera, tifo, peste (causati da batteri) e rabbia (causata da un virus). Agli inizi del XX secolo, statistici del calibro di Karl Pearson, Major Greenwood e Udny Yule furono attivamente coinvolti nelle discussioni sulla valutazione di questi vaccini sul campo.
Negli anni Venti comparvero nuovi vaccini, tra cui quelli contro pertosse, difterite, tetano e il bacillo di Calmette–Guérin contro la tubercolosi. Negli anni Trenta furono sviluppati i vaccini contro febbre gialla, influenza e rickettsie.
Dopo la Seconda guerra mondiale, l'introduzione delle colture cellulari, che consentivano la crescita dei virus, rese possibile la produzione dei vaccini contro poliomielite, morbillo, parotite, rosolia, varicella e adenovirus, tra gli altri. Successive innovazioni tecnologiche permisero la sostituzione di vecchie generazioni di vaccini con nuove, più efficaci, contro specifici agenti patogeni. La ricerca è tuttora in corso per vaccini contro malaria, HIV e molti altri agenti infettivi capaci di eludere non solo gli sforzi dei ricercatori, ma anche la risposta immunitaria naturale.

Alcuni vaccini risultano altamente efficaci, con effetti protettivi osservabili anche senza analisi statistiche sofisticate. Altri invece hanno un'efficacia minore, rendendo più complessi la progettazione degli studi e l'analisi statistica.
L'inferenza statistica ha compiuto grandi progressi nel XX secolo e continua a svilupparsi nel XXI. Lo sviluppo della statistica, del design degli studi clinici e dei metodi epidemiologici nel XX secolo ha avuto un parallelo nelle ricerche sui vaccini. Storicamente, l'attenzione si è focalizzata sulla valutazione degli effetti protettivi diretti dell'immunizzazione sugli individui vaccinati, sui quali si concentra l'analisi in questo contesto [30].
Allo scopo di stimare gli effetti protettivi diretti della vaccinazione, nel 1915 Greenwood e Yule [18] hanno indicato tre condizioni necessarie per un'inferenza valida:

1. Le persone devono essere, sotto tutti gli aspetti rilevanti, simili.

2. L'esposizione effettiva alla malattia deve essere identica per i soggetti vaccinati e per quelli non vaccinati.

3. Il modo in cui si stabilisce se una persona è stata vaccinata non deve influenzare il modo in cui si stabilisce se quella persona ha avuto la malattia e viceversa (in altre parole, i due criteri di accertamento devono essere indipendenti).

I trial randomizzati in doppio cieco sono progettati per garantire che tali criteri siano soddisfatti. Se queste condizioni sono rispettate e i gruppi di vaccinati e non vaccinati risultano ugualmente esposti all'infezione, allora eventuali differenze di rischio tra i due bracci sono verosimilmente attribuibili agli effetti biologici del vaccino.
L'obiettivo d'efficacia (endpoint) primario di un trial vaccinale viene introdotto nella sezione 3.2 attraverso un'appropriata definizione della misura del rischio di infezione. Nella sezione 3.3 sono presentati i principali metodi statistici di riferimento, sia frequentisti sia bayesiani, ampiamente discussi in letteratura, che sono stati rivisitati da Ewell [11] e impiegati nel trial sul vaccino anti-COVID-19 sponsorizzato da Pfizer/BioNTech nel 2020, come riportato da Polack [12] e Thomas [50]. Infine, nella sezione 3.4 viene proposto un nuovo approccio bayesiano, che verrà confrontato con i metodi standard nel capitolo 4.

3.2 Endpoint primario di un vaccino

3.2.1 Effetti vaccinali rispetto alla suscettibilità e alla malattia

I trial clinici vaccinali sono progettati per valutare quanto bene un vaccino protegga i soggetti ai quali è stato somministrato. Dunque, l'endpoint primario dello studio è una misura di quanto la vaccinazione sia efficace nel prevenire l'infezione o la malattia. A seconda che la protezione sia dalla sola infezione o dalla malattia vera e propria, si parla di *efficacia del vaccino rispetto alla suscettibilità* o di *efficacia del vaccino rispetto alla malattia* rispettivamente. Secondo quest'ultima definizione, un partecipante è considerato infetto e, quindi contribuisce come caso, in presenza di sintomi compatibili con la malattia, oltre a una conferma virologica (presenza del virus nell'organismo) o sierologica (anticorpi) con test positivo. Di conseguenza, l'infezione asintomatica non viene conteggiata come caso nell'analisi di efficacia rispetto alla malattia. Tuttavia, molte volte, la distinzione tra i due tipi di efficacia viene resa evidente semplicemente dalla definizione di caso utilizzata nello studio e dal metodo di accertamento.
Ad esempio, il trial clinico di Pfizer-BioNTech si è focalizzato su due endpoint primari, che riguardavano entambi la prevenzione della malattia (Covid-19 sintomatico), non la prevenzione dell'infezione in sé. Il primo endpoint primario era l'efficacia del vaccino BNT162b2 contro il Covid-19 confermato, con insorgenza almeno 7 giorni dopo la seconda dose, nei partecipanti che non avevano evidenza sierologica o virologica di infezione da SARS-CoV-2 fino a 7 giorni dopo la seconda dose (21 giorni tra una dose e l'altra); il secondo endpoint primario era l'efficacia contro il Covid-19 nei partecipanti con e senza evidenza di precedente infezione.

3.2.2 Tassi di incidenza come misure di rischio

Operativamente, l'*efficacia del vaccino* è una misura della riduzione percentuale del rischio relativo di sviluppare la malattia (o contrarre l'infezione) nel gruppo di trattamento sperimentale rispetto al gruppo di controllo (al quale viene somministrato un placebo oppure un altro vaccino). Una misura di rischio di uso comune in Epidemiologia è il *tasso di incidenza* (*incidence rate*), definito come il rapporto tra il numero di nuovi casi di malattia durante un determinato periodo e il tempo di osservazione (sorveglianza) di ciascun individuo, sommato su tutti gli individui. Per comprendere il motivo di questa scelta, è utile descrivere formalmente il processo di reclutamento dei partecipanti e il

relativo meccanismo di censura.

Sia D la durata dello studio in tempo reale (misurata in anni), sia R_i il tempo aleatorio di reclutamento (in anni) trascorso dall'inizio dello studio (e comunque precedente alla sua conclusione) e sia T_i il tempo aleatorio (in anni) dal reclutamento fino all'infezione dell'i-esimo paziente. Allora, se il paziente non abbandona lo studio prima della sua conclusione, il tempo aleatorio di censura è dato da $C_i = D - R_i$, dove D è una quantità fissa. Ogni paziente contribuisce al tempo totale di sorveglianza con una durata pari al minimo tra T_i e C_i. Analogamente, l'i-esimo paziente incrementa il numero totale di infezioni di 1 se $T_i < C_i$ e di 0 altrimenti. La Figura 3.1 ne fornisce una rappresentazione grafica. Il tempo totale di sorveglianza

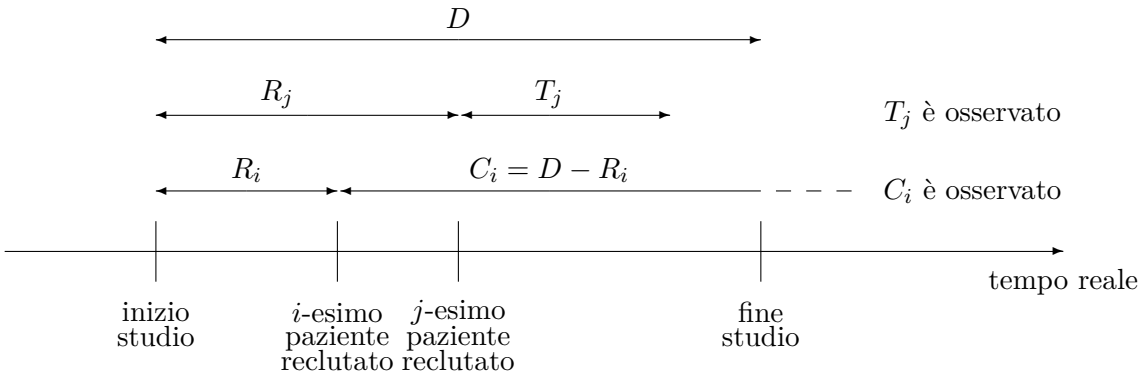


Figura 3.1: Reclutamento e censura casuali: l'i-esimo paziente è censurato, il j-esimo paziente non lo è.

$$S = \sum_{i=1}^{n} \min\{T_i, C_i\} \tag{3.1}$$

e il numero di infezioni

$$X = \sum_{i=1}^{n} \mathbb{1}_{\{T_i < C_i\}} \tag{3.2}$$

sono quindi somme di tante variabili aleatorie i.i.d. quanti sono i pazienti, dove $\mathbb{1}_{\{T_i<C_i\}}$ è la funzione indicatrice della disuguaglianza specificata e $\min\{T_i, C_i\}$ denota il tempo di sorveglianza individuale del soggetto i. I dati risultano quindi composti da due statistiche per gruppo:

- S_v = tempo di sorveglianza del gruppo vaccinato [persona-anni],

- S_c = tempo di sorveglianza del gruppo di controllo [persona-anni],
- X_v = numero di infezioni nel gruppo vaccinato,
- X_c = numero di infezioni nel gruppo di controllo.

Gli esempi 2.1 e 2.2 mostrano che, se i tempi di infezione di ciascun gruppo sono variabili aleatorie esponenziali i.i.d di parametro λ_v per il gruppo vaccinato e di parametro λ_c per il gruppo di controllo, allora le quattro statistiche sono sufficienti per stimare i due parametri e i rapporti X_v/S_v e X_c/S_c, ovvero i tassi di incidenza osservati per braccio, sono stimatori di massima verosimiglianza di λ_v e λ_c rispettivamente. Appare quindi naturale interpretare i parametri λ_v e λ_c delle due distribuzioni esponenziali come i tassi di incidenza veri della popolazione dei vaccinati e della popolazione dei non vaccinati rispettivamente, dalle quali sono stati estratti i relativi gruppi. Questi parametri riflettono l'intensità dei processi puntuali che descrivono le nuove infezioni in ciascun gruppo in una situazione stazionaria (cioè senza picchi rilevanti di contagio). Più precisamente, per ciascun gruppo un processo di nuove infezioni si ottiene concatenando, uno dopo l'altro, tutti i periodi di tempo durante i quali i partecipanti sono stati esposti alla possibilità di contrarre la malattia, fino al verificarsi del primo tra i seguenti quattro eventi: insorgenza della malattia, decesso, abbandono o fine dello studio. La somma delle durate di tutti i periodi costituisce il tempo totale di sorveglianza, mentre ciascun partecipante che sperimenta l'insorgenza della malattia contribuisce con un evento al processo di infezione. Gli altri eventi contribuiscono soltanto al tempo di sorveglianza, poiché si assume che essi corrispondano a osservazioni censurate in modo non informativo (solo per comodità C_i è stata definita come una forma di censura non informativa dovuta alla fine dello studio).
Una misura del rischio relativo comunemente utilizzata per confrontare i due processi di infezione è il *rapporto dei tassi di incidenza* (IRR), definito come

$$\text{IRR} = \frac{\lambda_v}{\lambda_c},$$

sulla base del quale l'efficacia vaccinale (VE) è definita come

$$\text{VE} = 1 - \text{IRR} = 1 - \frac{\lambda_v}{\lambda_c}. \tag{3.3}$$

La versione percentuale di tale quantità può essere interpretata, nel caso usuale in cui $\lambda_v < \lambda_c$, come la percentuale media di infezioni evitate (cioè la percentuale di soggetti vaccinati non infettati che si sarebbero infettati in assenza di vaccinazione).

3.3 Metodi standard per l'analisi di efficacia

In letteratura esistono diversi metodi consolidati per stimare l'efficacia di un vaccino (ed eventualmente condurre un test di ipotesi). Nel suo articolo [11], Ewell ha presentato una rassegna di diversi metodi per il calcolo degli intervalli statistici per l'efficacia vaccinale (VE), tra cui il metodo asintotico della massima verosimiglianza ("metodo B"), un approccio bayesiano con prior coniugate ("metodo A") e il metodo esatto di Clopper-Pearson basato sulla distribuzione binomiale ("metodo C"). Un approccio ibrido, che combina questi due ultimi metodi, è il cosiddetto *modello beta-binomiale*, ovvero un metodo bayesiano esatto condizionato al numero totale di casi: in questo modello, alla distribuzione binomiale del numero di pazienti infettati nel gruppo di trattamento, condizionata al numero totale di casi e ai tempi di sorveglianza di ciascun braccio, viene associata una prior coniugata Beta in maniera analoga al già citato metodo A, al fine di ottenere intervalli di credibilità per VE.
In questo contesto, ci si pone l'obbiettivo di confrontare l'approccio bayesiano ibrido e i due metodi frequentisti con un nuovo approccio bayesiano in cui si tiene conto anche dell'incertezza dei tempi di sorveglianza. A tale scopo, la funzione di verosimiglianza dei tassi di incidenza incogniti dei due gruppi, λ_v e λ_c, si può scrivere nella sua forma duale come la densità di probabilità congiunta delle due statistiche sufficienti (X_v, S_v) e (X_c, S_c), avente i due parametri come valori condizionanti:

$$f(s_v, s_c, x_v, x_c|\lambda_v, \lambda_c) = f_{S_v,S_c}(s_v, s_c|\lambda_v, \lambda_c) \times f_{X_v|S_v}(x_v|s_v, \lambda_v) \times f_{X_c|S_c}(x_c|s_c, \lambda_c). \quad (3.4)$$

I metodi bayesiani discussi dipendono dalle assunzioni modellistiche sui diversi fattori della verosimiglianza (3.4).

3.3.1 Stima asintotica dell'efficacia vaccinale

Il metodo della massima verosimiglianza (ML) è un approccio statistico che permette di derivare un intervallo di confidenza asintotico per VE. Infatti, poichè la statistiche X_v/S_v e X_c/S_c sono gli stimatori di massima verosimiglianza dei tassi di incidenza λ_v e λ_c rispettivamente, dalla proprietà di invarianza 2.5 segue che lo stimatore di massima verosimiglianza del rapporto dei tassi di incidenza IRR risulta essere

$$\widehat{\mathrm{IRR}} = \frac{X_v/S_v}{X_c/S_c}. \quad (3.5)$$

Dalle equazioni (2.2) e (2.5) si deduce che, per campioni sufficientemente grandi,

$$\log\left(\widehat{\mathrm{IRR}}\right) \sim \mathcal{N}\left(\log\left(\frac{\lambda_v}{\lambda_c}\right), \frac{1}{n_v p_v} + \frac{1}{n_v p_v}\right), \tag{3.6}$$

dove

$$p_v = \mathbb{P}(T_v \leq C)$$

e

$$p_c = \mathbb{P}(T_c \leq C)$$

sono le probabilità che un soggetto si infetti entro la fine dello studio nel gruppo vaccinato e di controllo rispettivamente. Quindi, la definizione (3.2) del numero di casi, applicata a ciascun gruppo, suggerisce che le statistiche X_v e X_c hanno entrambe una distribuzione binomiale:

$$X_v \sim \mathrm{Bin}(n_v, p_v) \quad \text{e} \quad X_c \sim \mathrm{Bin}(n_c, p_c),$$

dove n_v è il numero di pazienti vaccinati e n_c il numero di pazienti non vaccinati. Approssimando la varianza della distribuzione (3.6) con gli stimatori di massima verosimiglianza dei relativi parametri binomiali e sostituendola nella disuguaglianza (2.6), si ricava l'intervallo di confidenza asintotico con copertura $1-\alpha$ per il rapporto IRR in scala logaritmica:

$$\log\left(\frac{X_v\, S_c}{X_c\, S_v}\right) - z_{1-\alpha/2}\sqrt{\frac{1}{X_v} + \frac{1}{X_c}} \leq \log(\mathrm{IRR}) \leq \log\left(\frac{X_v\, S_c}{X_c\, S_v}\right) + z_{1-\alpha/2}\sqrt{\frac{1}{X_v} + \frac{1}{X_c}}. \tag{3.7}$$

Infine, applicando la trasformazione esponenziale e la definizione (3.3) di VE all'intervallo (3.7), si ottiene il corrispondente intervallo di confidenza approssimato per VE:

$$1-\exp\left(\log\left(\frac{X_v\, S_c}{X_c\, S_v}\right) + z_{1-\alpha/2}\sqrt{\frac{1}{X_v} + \frac{1}{X_c}}\right) \;\leq\; \mathrm{VE} \;\leq\; 1-\exp\left(\log\left(\frac{X_v\, S_c}{X_c\, S_v}\right) - z_{1-\alpha/2}\sqrt{\frac{1}{X_v} + \frac{1}{X_c}}\right).$$

Per evitare che il rapporto osservato tra i tassi di incidenza risulti indefinito qualora non si osservino eventi in uno dei gruppi (cioè quando il numero di casi è pari a zero), si aggiunge 0.5 sia al numero di soggetti infetti che agli anni-persona a rischio in ciascun gruppo. Questo accorgimento consente di calcolare l'intervallo di confidenza asintotico anche in presenza di conteggi nulli.

3.3.2 Verosimiglianza binomiale condizionata al numero totale di casi

In Epidemiologia, un'assunzione ampiamente diffusa è che il processo puntuale di infezione risultante dall'operazione di concatenazione descritta in precedenza è un processo di Poisson omogeneo, sebbene ciò possa costituire solo una buona approssimazione. Più precisamente, è consuetudine assumere che il numero di eventi osservati per il soggetto i segua una distribuzione di Poisson con parametro λS_i, dove λ è il tasso di incidenza e S_i il tempo di sorveglianza individuale. Questa ipotesi, a prima vista, sembra incompatibile con la realtà: in un trial ogni individuo può sviluppare al più un evento osservabile (prima infezione), mentre il modello di Poisson ammette in linea teorica un numero arbitrario di eventi. La giustificazione si basa sull'ipotesi di eventi rari. Per valori piccoli di λS_i, infatti:

$$\Pr(X_i = 0) \approx 1 - \lambda S_i, \qquad \Pr(X_i = 1) \approx \lambda S_i, \qquad \Pr(X_i \geq 2) \approx \tfrac{(\lambda S_i)^2}{2} \ll 1.$$

In tali condizioni, la probabilità che un soggetto contribuisca con più di un evento è trascurabile e la variabile di Poisson si comporta di fatto come una variabile di Bernoulli con parametro $1 - e^{-\lambda S_i}$ (ossia la probabilità di osservare almeno un evento entro S_i), in quanto il tempo T_i di prima infezione è esponenziale di parametro λ per ipotesi. L'adozione del modello di Poisson porta a una semplificazione analitica rilevante: i conteggi individuali si sommano esattamente in una variabile Poisson$(\lambda \sum_i S_i)$. Ciò consente di lavorare direttamente con il numero totale X di casi e il totale S degli anni-persona a rischio, evitando la complessità della somma di Bernoulli con parametri diversi (la cosiddetta *Poisson-binomiale*). Con questa assunzione, le funzioni $f_{X_v|S_v}$ e $f_{X_c|S_c}$ che compaiono nella verosimiglianza (3.4) sono densità di Poisson con parametri $\lambda_v S_v$ e $\lambda_c S_c$ rispettivamente. Basta poi assegnare delle prior Gamma coniugate ai tassi λ_v e λ_c per ottenere il già citato metodo A di Ewell.
Tuttavia, dal punto di vista informativo, risulta del tutto equivalente utilizzare il numero totale di infezioni osservate $x_v + x_c$ in luogo del numero x_c di casi osservati nel gruppo di controllo, in quanto vale l'uguaglianza

$$\begin{aligned} f_{X_v|S_v}(x_v|s_v, \lambda_v) \times f_{X_v|S_v}(x_c|s_c, \lambda_c) = f_{X_v+X_c|S_v,S_c}(x_v + x_c|s_v, s_c, \lambda_v, \lambda_c) \\ \times f_{X_v|X_v+X_c,S_v,S_c}(x_v|x_v + x_c, s_v, s_c, \lambda_v, \lambda_c), \end{aligned}$$

dove si verifica facilmente che il primo fattore del membro di destra è una densità di Poisson con parametro $\lambda_v S_v + \lambda_c S_c$, ottenuta dalla sovrapposizione di due processi di

Poisson indipendenti, mentre il secondo è una densità binomiale condizionata al numero totale di casi, cioè

$$f_{X_v|X_v+X_c,S_v,S_c}(x_v|x_v+x_c,s_v,s_c,\lambda_v,\lambda_c)=\binom{x_v+x_c}{x_v}\left(\frac{s_v\lambda_v}{s_v\lambda_v+s_c\lambda_c}\right)^{x_v}\left(1-\frac{s_v\lambda_v}{s_v\lambda_v+s_c\lambda_c}\right)^{x_c}. \tag{3.8}$$

La probabilità di successo

$$\theta=\frac{s_v\lambda_v}{s_v\lambda_v+s_c\lambda_c}=\frac{s_v(1-\mathrm{VE})}{s_v(1-\mathrm{VE})+s_c} \tag{3.9}$$

della distribuzione binomiale condizionata (3.8) rappresenta la probabilità che un individuo infetto appartenga al gruppo vaccinato. Condizionatamente ai tempi di sorveglianza, la relazione (3.9) suggerisce che conoscere VE equivale a conoscere la probabilità θ. Di conseguenza, trattando i tempi di sorveglianza osservati $S_v=s_v$ e $S_c=s_c$ come costanti, si può stimare θ e poi ottenere la corrispondente stima di VE tramite l'equazione (3.9).

3.3.2.1 Modello bayesiano condizionato

Come visto nella sezione 2.6, un metodo bayesiano richiede di specificare una funzione di verosimiglianza dipendente dal parametro di interesse e una distribuzione a priori che riflette l'informazione soggettiva sul parametro. In particolare, l'approccio bayesiano per la stima di θ consiste nell'assumere che il termine $f_{S_v,S_c}(\cdot)\times f_{X_v+X_c|S_v,S_c}(\cdot)$ della verosimiglianza (3.4) non dipende dal parametro VE, utilizzando di fatto la distribuzione binomiale condizionata (3.8) come funzione di verosimiglianza. Scegliendo una distribuzione a priori appartenente alla famiglia Beta(a,b), si ottiene un modello bayesiano coniugato, condizionato al numero totale di casi e ai tempi di sorveglianza di ciascun gruppo, che corrisponde al già citato modello beta-binomiale.
La selezione dei parametri a e b ha un duplice intento: avere una distribuzione a priori poco informativa (alta varianza) ed essere conservativi (media a priori sbilanciata a favore di una bassa efficacia del vaccino), lasciando che siano i dati a guidare la distribuzione a posteriori verso il vero ed eventualmente alto valore di VE (basso valore di θ). Dunque, poiché la varianza di una Beta è inversamente proporzionale ai suoi parametri, i valori di a e b dovrebbero essere sufficientemente piccoli per ottenere una prior poco informativa. In letteratura è comune centrare la prior indotta su VE da θ attorno a una stima preliminare $\widehat{\mathrm{VE}}\in(0{,}1)$, tipicamente bassa, ad esempio $\widehat{\mathrm{VE}}=0.30$, in modo da non distorcere il processo di stima verso valori troppo ottimistici. A tale scopo, si può

ancorare la prior di θ a una stima preliminare $\hat{\theta}$ della probabilità di infezione ottenuta sostituendo $\widehat{\text{VE}}$ nell'equazione (3.9). Però, in questo modo, la prior dipenderebbe dai dati e non rifletterebbe una conoscenza a priori rispetto al campionamento. Per superare una tale problematica, si può assumere a priori che i tempi di sorveglianza siano uguali ($s_v = s_c$), ottenendo così una stima preliminare di θ definita come

$$\hat{\theta} = \frac{(1 - \widehat{VE})}{2 - \widehat{VE}} \in \left(0, \frac{1}{2}\right). \tag{3.10}$$

Osservando che

$$\mathbb{E}[\theta] = \frac{a}{a+b},$$

è naturale ancorare la prior alla stima $\hat{\theta}$ ponendo

$$\hat{\theta} = \frac{a}{a+b}$$

e ricavando il primo parametro in funzione del secondo come

$$a = \frac{\hat{\theta}}{1-\hat{\theta}} b \in (0, b). \tag{3.11}$$

Per massimizzare la varianza, il parametro b andrebbe scelto il più piccolo possibile. Tuttavia, per $b < 1$ dall'equazione (3.11) si avrebbe $a < 1$ e la densità a priori risulterebbe bimodale. La bimodalità è poco conveniente perché, nel caso specifico, renderebbe la totale efficacia del vaccino (VE $= 1$) e la totale inefficacia (VE $= 0$) altamente verosimili, favorendo a priori sia un'elevata che una bassa efficacia del vaccino. Di conseguenza, per massimizzare la varianza ed evitare allo stesso tempo la bimodalità, si pone $b = 1$ ottenendo in definitiva una distribuzione a priori della forma

$$\theta \sim \text{Beta}\left(\frac{\hat{\theta}}{1-\hat{\theta}},\ 1\right). \tag{3.12}$$

Poichè la famiglia Beta è coniugata con la famiglia binomiale, la distribuzione a posteriori di θ è

$$\text{Beta}\left(\frac{\hat{\theta}}{1-\hat{\theta}} + x_v,\ 1 + x_c\right). \tag{3.13}$$

3.3.2.2 Stime intervallari esatte dell'efficacia vaccinale

Una volta ottenuta la distribuzione a posteriori, è possibile calcolare il seguente intervallo di credibilità di livello $1 - \alpha$ relativo a θ:

$$\text{Beta}_{\alpha/2}\left(\frac{\hat{\theta}}{1-\hat{\theta}} + x_v,\ 1 + x_c\right) \ \leq\ \theta\ \leq\ \text{Beta}_{1-\alpha/2}\left(\frac{\hat{\theta}}{1-\hat{\theta}} + x_v,\ 1 + x_c\right), \tag{3.14}$$

dove $\text{Beta}_{\alpha/2}(n, m)$ indica il quantile corrispondente al livello di probabilità $\alpha/2$ (l'$\alpha/2$-quantile) della distribuzione $\text{Beta}(n, m)$.
Inoltre, applicando il metodo frequentista discusso nel paragrafo 2.5.1, la statistica X_v, che condizionatamente al numero totale di casi $X_v + X_c = x_v + x_c$ segue la distribuzione binomiale (3.8), si può utilizzare per ottenere l'intervallo di confidenza esatto di Clopper-Pearson con livello di copertura (almeno) $1 - \alpha$:

$$\text{Beta}_{\alpha/2}(x_v, 1 + x_c) \;\leq\; \theta \;\leq\; \text{Beta}_{1-\alpha/2}(1 + x_v, x_c). \tag{3.15}$$

Questo intervallo di confidenza, per valori moderati di x_v e x_c, risulta praticamente indistinguibile dall'intervallo di credibilità (3.14) ed evita qualunque riferimento a una stima a priori $\hat{\theta}$, in quanto deriva da una procedura frequentista che preserva almeno il livello nominale in modo esatto. L'intervallo di credibilità bayesiano e l'intervallo conservativo di Clopper–Pearson forniscono stime intervallari esatte del parametro θ, che si possono invertire tramite la formula (3.9) in modo da ottenere le corrispondenti stime intervallari esatte di VE:

$$\frac{(1 - H)\, s_v - H\, s_c}{(1 - H)\, s_v} \;\leq\; \text{VE} \;\leq\; \frac{(1 - L)\, s_v - L\, s_c}{(1 - L)\, s_v}, \tag{3.16}$$

dove L e H rappresentano rispettivamente l'estremo sinistro e destro degli intervalli nelle equazioni (3.14) e (3.15). Anche in questo caso, per valori moderati di x_v e x_c, i due intervalli risultano sostanzialmente equivalenti. Tuttavia, è importante sottolineare che tutti questi intervalli sono stati ricavati *senza alcuna modellizzazione del processo di reclutamento*. L'obiettivo della presente analisi è il superamento di tale limitazione, al fine di sfruttare al meglio le informazioni contenute nei dati e, in particolare, quelle relative al processo di reclutamento, come verrà illustrato nella prossima sezione.

3.4 Approccio bayesiano con verosimiglianza completa

3.4.1 Motivazione del modello proposto

A onor del vero, assumere che la distribuzione congiunta $f_{S_v,S_c}(\cdot)$ dei tempi di sorveglianza totali, che compare nella definizione della funzione di verosimiglianza (3.4), sia indipendente da VE è soltanto una buona approssimazione. Infatti, come si è visto nel Capitolo 2, i vettori aleatori (S_v, X_v) e (S_c, X_c) sono statistiche sufficienti per i tassi di incidenza λ_v e λ_c rispettivamente. Di conseguenza, le variabili aleatorie S_v e S_c sono

statistiche informative e, in quanto tali, contribuiscono anch'esse a informare l'efficacia vaccinale.
Alla luce di queste considerazioni, nel seguito viene proposto un approccio bayesiano che tiene conto della natura aleatoria dei tempi di sorveglianza S_v e S_c, anziché trattarli semplicemente come costanti che definiscono i parametri dei processi infettivi di tipo Poisson. In particolare, si ricava una distribuzione asintotica congiunta delle quattro statistiche S_v, S_c, X_v, X_c, ovvero si deriva il comportamento asintotico delle densità che definiscono la funzione di verosimiglianza (3.4).

3.4.2 Verosimiglianza normale bivariata

Poiché il gruppo vaccinato e quello di controllo generano osservazioni indipendenti, la distribuzione congiunta di (S_v, X_v) (risp. (S_c, X_c)) viene trattata in questa paragrafo utilizzando la notazione unificata (S, X), ossia omettendo i pedici, poiché i medesimi risultati valgono sia per il gruppo vaccinato sia per il gruppo di controllo.

Teorema 3.1. *Sia D la durata del trial vaccinale (costante) e siano T_i il tempo aleatorio dal reclutamento all'infezione, R_i il tempo aleatorio di reclutamento e $C_i = D - R_i$ il tempo aleatorio di censura relativi all' i-esimo paziente, su un totale di n pazienti. Inoltre, si supponga che valgano le seguenti condizioni:*

1. *i tempi all'infezione $T_1, \dots, T_n \sim T$ formano un campione casuale estratto dalla distribuzione esponenziale di parametro λ della variabile aleatoria T;*

2. *i tempi di censura $C_1, \dots, C_n \sim C$ formano un campione casuale estratto dalla distribuzione della variabile aleatoria C e sono indipendenti dai tempi di infezione.*

Allora, usando le definizioni (3.1) e (3.2), per $n \to \infty$ vale la seguente convergenza in distribuzione a una varaibile aleatoria normale bivariata:

$$\frac{1}{\sqrt{n}} \begin{pmatrix} S - nI_1 \\ X - n\lambda I_1 \end{pmatrix} \longrightarrow \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2I_2 - I_1^2 & \lambda I_2 - \lambda I_1^2 \\ \lambda I_2 - \lambda I_1^2 & \lambda I_1 - \lambda^2 I_1^2 \end{pmatrix} \right), \tag{3.17}$$

dove

$$I_1 = \int_0^\infty e^{-\lambda t} \, \mathrm{P}(C > t) \, dt, \tag{3.18}$$

$$I_2 = \int_0^\infty t \, e^{-\lambda t} \, \mathrm{P}(C > t) \, dt. \tag{3.19}$$

Dimostrazione. La dimostrazione è una conseguenza del teorema limite centrale multivariato 2.3, in base al quale

$$\frac{1}{\sqrt{n}}\begin{pmatrix} S - n\,\mathrm{E}[\min(T,C)] \\ X - n\,\mathrm{E}\Big[\mathbb{1}_{\{T<C\}}\Big] \end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} \mathrm{Var}(\min(T,C)) & \mathrm{Cov}\Big(\min(T,C), \mathbb{1}_{\{T<C\}}\Big) \\ \mathrm{Cov}\Big(\min(T,C), \mathbb{1}_{\{T<C\}}\Big) & \mathrm{Var}\Big(\mathbb{1}_{\{T<C\}}\Big) \end{pmatrix}\right).$$

Ora, poiché sia T che C sono quasi certamente positivi e indipendenti, si ha

$$\begin{aligned}
\mathbb{E}[\min(T,C)] &= \int_0^\infty \mathrm{P}(\min(T,C) > t)\,dt \\
&= \int_0^\infty \mathrm{P}(T>t)\ \mathrm{P}(C>t)\,dt \\
&= \int_0^\infty e^{-\lambda t}\ \mathrm{P}(C>t)\,dt = I_1, \\
\mathbb{E}\Big[\min(T,C)^2\Big] &= \int_0^\infty \mathrm{P}(\min(T,C)^2 > t)\,dt \\
&= \int_0^\infty \mathrm{P}(T>\sqrt{t})\ \mathrm{P}(C>\sqrt{t})\,dt \\
&= 2\int_0^\infty t\ e^{-\lambda t}\ \mathrm{P}(C>t)\,dt = 2I_2.
\end{aligned}$$

La varianza della variabile $\min(T,C)$ si ottiene quindi come

$$\begin{aligned}
\mathrm{Var}(\min(T,C)) &= \mathbb{E}\Big[\min(T,C)^2\Big] - \Big(\mathbb{E}[\min(T,C)]\Big)^2 \\
&= 2I_2 - I_1^2.
\end{aligned}$$

Inoltre, essendo $\mathbb{1}_{\{T<C\}}$ una funzione indicatrice, ovvero una variabile di Bernoulli con probabilità di successo p, si ha

$$p = \mathbb{E}\Big[\mathbb{1}_{\{T<C\}}\Big] = \mathbb{P}(T<C) = \int_0^\infty \lambda e^{-\lambda t}\ \mathbb{P}(C>t)\,dt = \lambda I_1. \tag{3.20}$$

Infine, poichè

$$\mathbb{E}\Big[\,\min(T,C)\cdot\mathbb{1}_{\{T<C\}}\Big] = \int_0^\infty t\,\lambda\,e^{-\lambda t}\,\mathbb{P}(C>t)\,dt = \lambda I_2,$$

la dimostrazione si conclude calcolando la covarianza

$$\begin{aligned}
\mathrm{Cov}\Big(\min(T,C), \mathbb{1}_{\{T<C\}}\Big) &= \mathbb{E}\Big[\min(T,C)\cdot\mathbb{1}_{\{T<C\}}\Big] \ - \ \mathbb{E}[\min(T,C)]\ \cdot\ \mathbb{E}\Big[\mathbb{1}_{\{T<C\}}\Big] \\
&= \lambda I_2 - \lambda I_1^2.
\end{aligned}$$

□

È importante sottolineare che il Teorema fornisce una distribuzione asintotica di S e X che dipende dal processo di reclutamento, ma soltanto attraverso i primi due momenti dei tempi di sorveglianza. Pertanto, in questo approccio è possibile considerare qualsiasi tipologia di reclutamento, inclusi trial multicentrici, reclutamenti con ingresso scaglionato dei pazienti e trial in cui il reclutamento viene interrotto prima della conclusione. Si noti inoltre che la convergenza della formula (3.6) può essere ottenuta come conseguenza della convergenza della formula (3.17) mediante l'applicazione del metodo delta multivariato. Questa osservazione ha diverse implicazioni importanti:

- il risultato del Teorema costituisce un'estensione del consueto teorema di approssimazione asintotica per lo stimatore di massima verosimiglianza;
- l'ottimalità del metodo di massima verosimiglianza, per il nostro modello regolare, fornisce un riferimento (ad esempio in termini di ampiezza degli intervalli di stima per VE, sia frequentisti che bayesiani) che non può essere superato asintoticamente; tuttavia, con i risultati più generali qui presentati, vi è la prospettiva di migliorare le prestazioni in campioni finiti, nello spirito delle asintotiche per piccoli campioni;
- mentre la varianza nella formula (3.6) coinvolge solo il primo momento dei tempi di sorveglianza, l'approccio bivariato del teorema 3.1 coinvolge sia il primo sia il secondo momento dei tempi di sorveglianza, ma non richiede ulteriori dettagli riguardo al reclutamento.

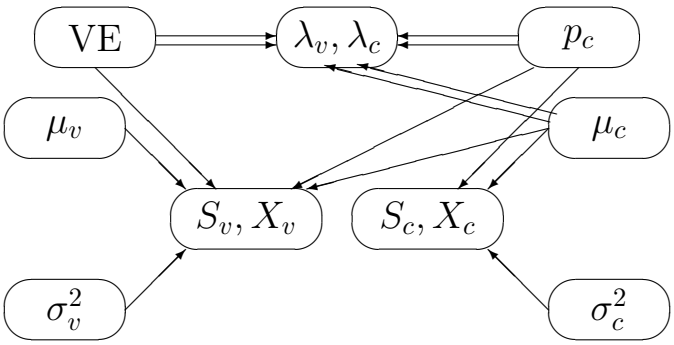


Figura 3.2: DAG per il modello bayesiano con verosimiglianza completa: le doppie frecce indicano la dipendenza funzionale di (λ_v, λ_c) da VE, p_c e μ_c.

3.4.3 Modello bayesiano completo

Applicando i risultati del teorema 3.1 al gruppo vaccinato e al gruppo di controllo, si può riscrivere la verosimiglianza (3.4) nella forma seguente:

$$\mathcal{L}(\mathrm{VE}, \pi_c, \mu_c, \sigma_c^2, \mu_v, \sigma_v^2; s_c, x_c, s_v, x_v) = f_{S_c,X_c}(s_c, x_c \mid \pi_c, \mu_c, \sigma_c^2) \cdot f_{S_v,X_v}(s_v, x_v \mid \mathrm{VE}, \pi_c, \mu_c, \mu_v, \sigma_v^2), \tag{3.21}$$

dove

$$\begin{aligned}
\mathrm{VE} &= 1 - \frac{\lambda_v}{\lambda_c}, \\
p_c &= \mathrm{P}(T_c < C) = \lambda_c I_{1,c}, \\
\mu_c &= I_{1,c} = \int_0^\infty e^{-\lambda_c t}\ \mathrm{P}(C > t)\, dt, \\
\sigma_c^2 &= 2I_{2,c} - I_{1,c}^2 = \int_0^\infty t\ e^{-\lambda_c t}\, \mathrm{P}(C > t)\, dt - I_{1,c}^2, \\
\mu_v &= I_{1,v} = \int_0^\infty e^{-\lambda_v t}\ \mathrm{P}(C > t)\, dt, \\
\sigma_v^2 &= 2I_{2,v} - I_{1,v}^2 = \int_0^\infty t\ e^{-\lambda_v t}\ \mathrm{P}(C > t)\, dt - I_{1,v}^2,
\end{aligned}$$

e $f_{S_c,X_c}(\cdot)$ e $f_{S_v,X_v}(\cdot)$ sono, rispettivamente, le densità delle seguenti distribuzioni asintotiche normali bivariate:

$$\begin{pmatrix} S_c \\ X_c \end{pmatrix} \sim \mathcal{N}_2 \left(\begin{pmatrix} n_c I_{1,c} \\ n_c \lambda_c I_{1,c} \end{pmatrix}, n_c \begin{pmatrix} 2I_{2,c} - I_{1,c}^2 & \lambda_c I_{2,c} - \lambda_c I_{1,c}^2 \\ \lambda_c I_{2,c} - \lambda_c I_{1,c}^2 & \lambda_c I_{1,c} - \lambda_c^2 I_{1,c}^2 \end{pmatrix} \right),$$

$$\begin{pmatrix} S_v \\ X_v \end{pmatrix} \sim \mathcal{N}_2 \left(\begin{pmatrix} n_v I_{1,v} \\ n_v \lambda_v I_{1,v} \end{pmatrix}, n_v \begin{pmatrix} 2I_{2,v} - I_{1,v}^2 & \lambda_v I_{2,v} - \lambda_v I_{1,v}^2 \\ \lambda_v I_{2,v} - \lambda_v I_{1,v}^2 & \lambda_v I_{1,v} - \lambda_v^2 I_{1,v}^2 \end{pmatrix} \right).$$

Si noti che la probabilità di infezione p_v di un paziente vaccinato non compare nella verosimigianza (3.21). Infatti, p_v è un parametro dipendente univocamente determinato dai parametri indipendenti VE, p_c, μ_v e μ_c mediante l'uguaglianza (3.20), cioè

$$p_v = p_c(1 - \mathrm{VE})\frac{\mu_\mathrm{v}}{\mu_\mathrm{c}}.$$

La ragione per parametrizzare la verosimiglianza in termini di $\mathrm{VE}, p_c, \mu_c, \sigma_c^2, \mu_v, \sigma_v^2$ (invece che, come fatto precedentemente, in termini di λ_v e λ_c) è, in primo luogo, quella di sottolineare che il parametro VE è il principale oggetto dell'inferenza, ma anche di

facilitare la convergenza dell'algoritmo MCMC utilizzato per approssimare il meccanismo di inferenza bayesiana a posteriori: con questa parametrizzazione, tutti i parametri eccetto VE hanno distribuzioni a supporto limitato e, in totale assenza di informazione, si potrebbe scegliere di assegnare a ciascuno di essi una prior uniforme. Utilizzando la stessa idea di trasformare i parametri in grandezze a supporto limitato, una prior di default per VE può essere quella indotta da

$$\frac{1-\mathrm{VE}}{2-\mathrm{VE}} \sim \mathrm{Beta}(1-\widehat{\mathrm{VE}}, 1), \tag{3.22}$$

che coincide con la prior (3.12) del modello bayesiano condizionato, dove i tempi di sorveglianza sono a priori assunti uguali.

Per gli altri parametri, in modo indipendente, si assume:

$$\begin{aligned} p_c &\sim \mathrm{Uniform}(0,1), \\ \mu_c &\sim \mathrm{Uniform}(0, D), \\ \sigma_c^2 &\sim \mathrm{Uniform}(0, D^2), \\ \mu_v &\sim \mathrm{Uniform}(0, D), \\ \sigma_v^2 &\sim \mathrm{Uniform}(0, D^2). \end{aligned}$$

Assegnando le suddette distribuzioni a priori ai parametri VE, $\pi_c, \mu_c, \sigma_c^2, \mu_v, \sigma_v^2$, si ottiene il modello bayesiano con verosimiglianza completa, rappresentato graficamente dal corrispondente grafo aciclico diretto (DAG) in Figura 3.2.
Infine, poiché le distribuzioni a priori non sono coniugate, un accorgimento per migliorare la convergenza e la stabilità dell'algoritmo MCMC consiste nel simulare (S_c, X_c) e (S_v, X_v) generando dapprima X_c (risp. X_v) dalla sua distribuzione binomiale esatta, e successivamente S_c (risp. S_v) dalla distribuzione normale condizionata indotta dalla distribuzione asintotica normale bivariata.

Capitolo 4

Risultati numerici

4.1 Presentazione dei risultati

In questo capitolo, l'approccio bayesiano con verosimiglianza completa (Full Bayesian approach) viene confrontato con gli altri metodi statistici standard in termini di precisione della stima dell'efficacia vaccinale, adottando diverse metriche di valutazione.
Il confronto è condotto con un approccio frequentista: si specifica il design sperimentale dello studio clinico e si simulano tanti studi con lo stesso design in scenari differenti, variando cioè il valore vero di VE, il numero di pazienti reclutati e la distribuzione dei tempi di reclutamento.
L'impostazione della simulazione è presentata nella sezione 4.2, mentre le metriche di valutazione impiegate e il confronto tra i modelli vengono discussi nella sezione 4.3. Si rimanda all'Appendice A per i dettagli implementativi.

4.2 Simulazione di studi clinici

4.2.1 Design dello studio simulato

La verosimiglianza del modello bayesiano proposto è condizionata alla durata D dello studio clinico. Appare quindi naturale adottare un design time-driven, trattando D come un parametro di controllo da scegliere in fase di progettazione dello studio. Per diversi motivi, come discusso nel capitolo 1, molti studi clinici invece adottano un design event-driven. Quest'ultimo tipo di design si rileva appropriato nel valutare le prestazioni del

modello proposto, in quanto esse sono strettamente legate al numero totale di casi osservati. Tuttavia, un design event-driven renderebbe la durata dello studio una variabile aleatoria e quindi le prior definite nel Capitolo 3 avrebbero un supporto dipendente dai dati, aspetto difficile da giustificare in un'impostazione bayesiana classica (nell'approccio Empirical Bayes un design di questo tipo sarebbe ammissibile). Per eludere questa problematica e altre complicazioni implementative legate al reclutamento che derivano da un design event-driven, si è scelto un design time-driven, variando però il numero di pazienti reclutati in funzione di un numero atteso di eventi prefissato.

4.2.2 Generazione degli scenari

Nello specifico, gli studi clinici vengono simulati con una durata fissa $D = 1$ e un tasso di infezione nel gruppo di controllo pari a $\lambda_c = 0.1$, valore dello stesso ordine di grandezza di quello stimato durante la pandemia di COVID-19. Tutti gli altri parametri liberi vengono variati al fine di valutare le prestazioni dell'approccio proposto in un'ampia gamma di scenari plausibili. La vera efficacia vaccinale (VE) varia tra 0.1 e 0.9, con incrementi di 0.2, per riflettere diversi regimi di efficacia. Di conseguenza, il tasso di infezione nel gruppo vaccinato è definito come $\lambda_v = (1 - \text{VE})\lambda_c$.
In ogni simulazione i partecipanti sono reclutati durante una finestra temporale pari a una frazione τ della durata D dello studio. In particolare, i dati sono generati a partire da due differenti processi di reclutamento:

- un *reclutamento uniforme*, in cui i soggetti entrano nello studio in tempi i.i.d. distribuiti come una variabile aleatoria R avente densità uniforme

  $$f_R(r) = \frac{1}{\tau D}, \quad r \in [0, \tau D],$$

  che riflette un processo di reclutamento a tasso costante nella finestra temporale $[0, \tau D]$;

- un *reclutamento beta riscalato*, in cui i soggetti entrano nello studio in tempi i.i.d. distribuiti come una variabile aleatoria R avente densità

  $$f_R(r) = 6\,\frac{r}{(\tau D)^2}\left(1 - \frac{r}{(\tau D)^2}\right), \quad r \in [0, \tau D],$$

  ovvero avente una distribuzione Beta(2,2) riscalata sull'intervallo $[0, \tau D]$, che riflette un processo di ingresso dei pazienti con tasso crescente nella fase iniziale e decrescente nella fase finale della finestra di reclutamento.

In entrambe le tipologie di reclutamento si assume $\tau = 0.75$, in modo tale che i partecipanti vengano reclutati in una finestra temporale pari ai tre quarti della durata dello studio. Questa scelta garantisce a ciascun soggetto un tempo minimo di esposizione all'infezione pari a 3 mesi.
Nel processo di generazione dei dati, per ciascun paziente i il tempo di reclutamento R_i è campionato dalla distribuzione di reclutamento scelta (Uniforme o Beta), mentre il tempo all'infezione T_i è campionato da una distribuzione esponenziale con tasso λ_v o λ_c a seconda che il paziente sia vaccinato o meno. I tempi di sorveglianza totali e i conteggi delle infezioni per gruppo sono quindi calcolati secondo le equazioni (3.1) e (3.2) rispettivamente.
Per ogni combinazione di VE e tipologia di reclutamento, la numerosità campionaria totale $n_c + n_v$ è calcolata in modo da ottenere un numero atteso di infezioni complessive $\mathbb{E}(X_c + X_v)$ tra i due gruppi che vari nell'insieme $\{40, 80, 160, 900\}$.

4.2.3 Randomizzazione a blocchi semplificata

Assumendo un rapporto di allocazione 1:1 tra gruppo vaccinato e gruppo di controllo, il numero totale di pazienti necessario è stimato mediante la seguente formula:

$$n_c = n_v = \frac{\mathrm{E}(X_c + X_v)}{\pi_c + \pi_v},$$

dove

$$\pi_c = 1 - \int_0^{\tau D} e^{-\lambda_c(D-r)} f_R(r)\, dr,$$
$$\pi_v = 1 - \int_0^{\tau D} e^{-\lambda_v(D-r)} f_R(r)\, dr.$$

Si osservi che, sebbene la simulazione implementi una randomizzazione bilanciata mediante allocazione alternata 1:1, essa riflette gli effetti desiderati di una randomizzazione a blocchi di dimensione fissa. Infatti, poichè i blocchi non sono associati ad alcuna covariata o tendenza temporale e il bilanciamento numerico tra i gruppi è perfetto, non è stato necessario modellare esplicitamente la struttura a blocchi nell'analisi bayesiana.

4.2.4 Campionamento dalla distribuzione a posteriori

Per ciascuno scenario di simulazione vengono generati 10,000 studi clinici. Un'approssimazione della distribuzione a posteriori di VE è ottenuta tramite campionamento

MCMC, utilizzando la piattaforma `JAGS` [38] integrata nel software `R` [41]. In particolare, per ogni ogni studio simulato vengono generate tre catene MCMC da 20,000 iterazioni ciascuna, con un burn-in di 2,000 iterazioni e un fattore di thinning pari a 20, in modo da ottenere circa 3,000 campioni a posteriori effettivamente indipendenti. Parallelamente, 3,000 campioni vengono generati dalla distribuzione a posteriori (3.13), che caratterizza il modello bayesiano binomiale condizionato (Conditional Bayesian approach).

4.3 Valutazione delle prestazioni

4.3.1 Metriche di valutazione

Le prestazioni del nuovo approccio bayesiano con verosimiglianza completa (FB) sono confrontate con quelle degli altri metodi standard adottando le seguenti metriche:

- *Copertura media degli intervalli di credibilità o confidenza al 95%* di VE per ciascun metodo, al fine di valutare la calibrazione frequentista e l'assenza di un'eccessiva influenza delle distribuzioni a priori;

- *Riduzione percentuale media dell'ampiezza degli intervalli al 95%* di VE ottenuta con l'approccio FB rispetto ai metodi concorrenti, come misura della maggior precisione degli intervalli;

- *Riduzione percentuale dell'errore quadratico medio* (MSE) dello stimatore puntuale di VE ottenuta con l'approccio FB (media a posteriori) rispetto agli altri metodi, come misura della qualità del compromesso tra precisione (varianza) e accuratezza (bias) dello stimatore. In quest'ottica, si considerano, come metriche ausiliarie, anche le corrispondenti riduzioni percentuali del bias quadratico e della varianza dello stimatore puntuale ottenute con l'approccio FB.

4.3.2 Confronto tra i modelli

I risultati numerici (incluse le metriche ausiliarie) dello studio simulativo sono riassunti nelle tabelle 4.1 e 4.3 per quanto riguarda il reclutamento uniforme e nelle tabelle 4.2 e 4.4 per quanto riguarda il reclutamento beta. Inoltre, per ciascun tipo di reclutamento, la riduzione percentuale dell'ampiezza degli intervalli e la riduzione percentuale dell'MSE sono rappresentati graficamente nelle figure 4.1 e 4.2 rispettivamente.

4.3.2.1 Confronto delle stime intervallari

Per quanto riguarda la copertura media degli intervalli di credibilità al 95%, il metodo bayesiano completo (FB) si avvicina costantemente al livello nominale in tutti gli scenari analizzati. Un comportamento simile si osserva per l'approccio bayesiano condizionato (CB). Al contrario, gli intervalli di confidenza esatti di Clopper–Pearson (CP) e gli intervalli di confidenza asintotici del metodo di massima verosimiglianza (ML) mostrano una copertura media costantemente superiore al livello nominale, confermando la natura conservativa di entrambe le procedure, in particolare quando il numero di eventi osservati è ridotto e le assunzioni asintotiche potrebbero non essere valide. Alla luce di questi risultati e delle piccole oscillazioni intorno al livello nominale, tutti i metodi considerati possono essere ritenuti ben calibrati in senso frequentista [19, 44].
Confrontando l'ampiezza degli intervalli al 95% di VE (intervalli di credibilità o di confidenza, a seconda che l'approccio sia bayesiano o frequentista), il metodo FB mostra un miglioramento uniforme rispetto ai metodi concorrenti in quasi tutti gli scenari considerati. In particolare, la riduzione dell'ampiezza rispetto agli altri approcci è più marcata quando il numero di infezioni osservato è piccolo o moderato e tende a ridursi progressivamente all'aumentare del numero di infezioni. È interessante notare che l'ampiezza dell'intervallo di credibilità FB risulta molto vicina a quella dell'intervallo asintotico ML quando il numero atteso di infezioni è almeno 900.
Poichè la distribuzione a priori dei due approcci bayesiani è la stessa, essi possono essere confrontati in maniera equa. A parità di numero atteso di infezioni, si osserva una riduzione percentuale media più elevata dell'ampiezza dell'intervallo di credibilità FB rispetto a quello CB per valori bassi di VE, mentre emerge un sostanziale accordo tra i due approcci quando $\mathbb{E}[X_v + X_c] = 900$, a prescindere dal valore di VE. Infatti, in quest'utlimo scenario, i valori negativi riportati nelle tabelle 4.1 e 4.2 relativi al confronto tra il metodo FB e il metodo CB sono di entità così ridotta da poter considerare i due approcci sostanzialmente equivalenti in termini di stima intervallare.
Confrontando il metodo FB con i due approcci frequentisti, si osserva, a parità di casi, una riduzione dell'ampiezza degli intervalli di entità simile per diversi valori di VE, in particolare quando VE è bassa o moderata. Tuttavia, una riduzione percentuale media significativamente maggiore è osservata per il metodo FB quando VE è pari a 0.9. Questo risultato è coerente con la tendenza, ampiamente documentata, degli intervalli di confidenza frequentisti a diventare eccessivamente conservativi per valori elevati di VE.

Non si osservano differenze rilevanti nella riduzione percentuale media dell'ampiezza degli intervalli tra i due processi di reclutamento considerati (uniforme e beta).

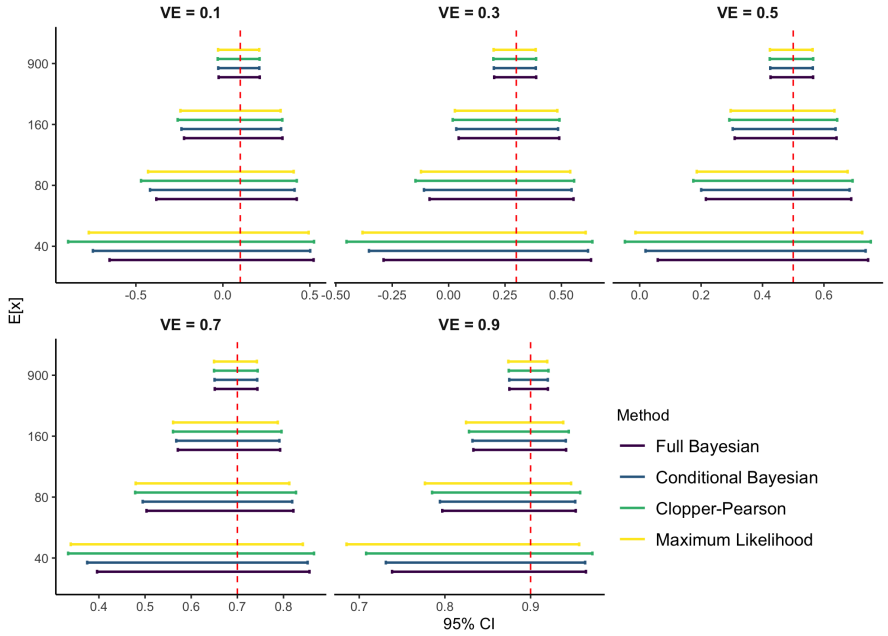
Tabella 4.1: Confronto tra l'approccio bayesiano con verosimiglianza completa e gli altri approcci in termini di stima intervallare: reclutamento uniforme.

VE	**R**	$E[\mathbf{X_c}+\mathbf{X_v}]$	$\mathbf{n_c}+\mathbf{n_v}$	**Coverage**				**Interval width % reduction**		
				FB	**CB**	**CP**	**ML**	**CB**	**CP**	**ML**
0.1	Uniform	40	697	0.948	0.949	0.963	0.955	5.519	16.081	6.931
0.1	Uniform	80	1393	0.952	0.951	0.964	0.953	2.693	9.618	3.495
0.1	Uniform	160	2786	0.946	0.946	0.956	0.949	1.199	5.927	1.721
0.1	Uniform	900	15668	0.946	0.947	0.951	0.948	-0.105	1.833	0.108
0.3	Uniform	40	777	0.952	0.952	0.968	0.958	4.846	14.959	6.907
0.3	Uniform	80	1553	0.949	0.950	0.963	0.955	2.324	9.102	3.489
0.3	Uniform	160	3105	0.948	0.949	0.958	0.952	1.136	5.759	1.795
0.3	Uniform	900	17465	0.951	0.951	0.955	0.952	-0.121	1.842	0.146
0.5	Uniform	40	879	0.950	0.951	0.966	0.957	4.090	14.000	7.256
0.5	Uniform	80	1757	0.950	0.952	0.964	0.955	1.973	8.687	3.671
0.5	Uniform	160	3514	0.952	0.951	0.961	0.953	0.916	5.575	1.859
0.5	Uniform	900	17465	0.951	0.951	0.955	0.952	-0.121	1.842	0.146
0.7	Uniform	40	1014	0.950	0.948	0.966	0.957	3.406	13.411	8.755
0.7	Uniform	80	2028	0.952	0.951	0.966	0.953	1.657	8.622	4.446
0.7	Uniform	160	4055	0.952	0.952	0.963	0.955	0.785	5.710	2.278
0.7	Uniform	900	22808	0.953	0.954	0.958	0.955	-0.005	2.176	0.417
0.9	Uniform	40	1202	0.957	0.955	0.977	0.960	2.568	14.525	17.801
0.9	Uniform	80	2403	0.953	0.953	0.971	0.956	1.280	9.998	9.168
0.9	Uniform	160	4806	0.949	0.948	0.962	0.949	0.626	7.074	4.714
0.9	Uniform	900	27030	0.949	0.949	0.955	0.949	-0.006	2.957	0.869

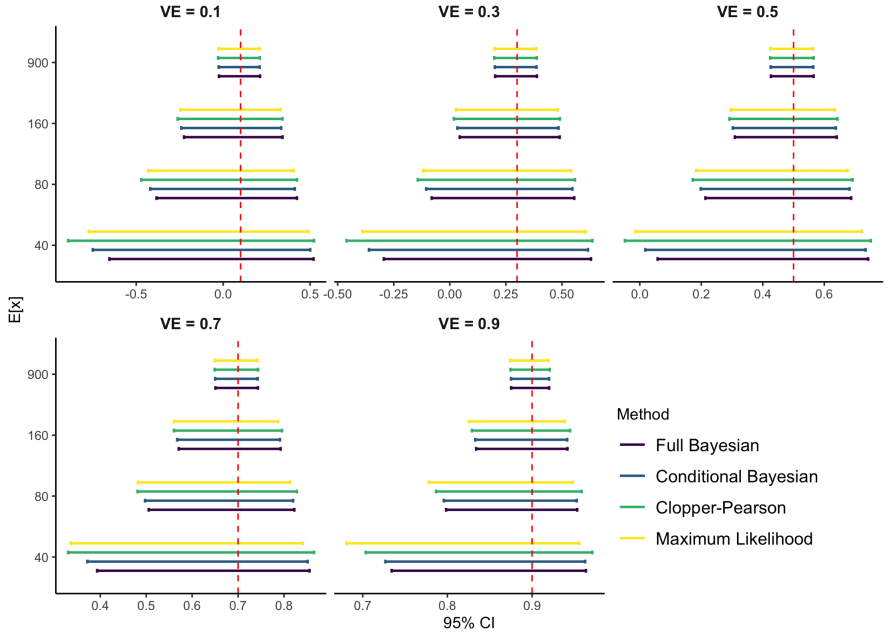
FB: Full Likelihood Bayesian, CB: Conditional Bayesian, CP: Clopper-Pearson frequentist, ML: Maximum Likelihood frequentist.

4.3.2.2 Confronto delle stime puntuali

Il vantaggio dell'approccio FB proposto è evidente anche nella stima puntuale della VE, come mostrato dalla riduzione percentuale media dell'errore quadratico medio (MSE) rispetto ai metodi concorrenti. In questa analisi, la media a posteriori è impiegata come stima puntuale di VE per gli approcci bayesiani, mentre per i metodi frequentisti si utilizza la stima di massima verosimiglianza ottenuta applicando la definizione (3.3) di VE all'equazione (3.5).



(a) Reclutamento uniforme



(b) Reclutamento beta

Figura 4.1: Confronto degli intervalli statistici per VE con una probabilità di copertura/credibilità del 95%.

Tabella 4.2: Confronto tra l'approccio bayesiano con verosimiglianza completa e gli altri approcci in termini di stima intervallare: reclutamento beta riscalato.

VE	**R**	$\mathbf{E[X_c + X_v]}$	$\mathbf{n_c + n_v}$	**Coverage**				**Interval width % reduction**		
				FB	**CB**	**CP**	**ML**	**CB**	**CP**	**ML**
0.1	Beta	40	696	0.950	0.949	0.964	0.956	5.594	16.077	6.941
0.1	Beta	80	1391	0.951	0.951	0.963	0.955	2.730	9.647	3.525
0.1	Beta	160	2782	0.950	0.952	0.959	0.953	1.279	5.963	1.756
0.1	Beta	900	15647	0.948	0.946	0.952	0.949	-0.151	1.817	0.091
0.3	Beta	40	776	0.952	0.952	0.967	0.959	4.951	15.048	6.988
0.3	Beta	80	1551	0.952	0.952	0.964	0.955	2.371	9.122	3.519
0.3	Beta	160	3101	0.950	0.952	0.960	0.954	1.067	5.714	1.751
0.3	Beta	900	17443	0.952	0.951	0.956	0.952	-0.254	1.711	0.012
0.5	Beta	40	878	0.952	0.951	0.967	0.957	4.082	13.925	7.198
0.5	Beta	80	1755	0.951	0.950	0.963	0.955	1.983	8.705	3.690
0.5	Beta	160	3510	0.948	0.948	0.957	0.950	0.919	5.614	1.900
0.5	Beta	900	19740	0.948	0.948	0.952	0.949	-0.206	1.817	0.122
0.7	Beta	40	1013	0.952	0.950	0.968	0.958	3.369	13.417	8.707
0.7	Beta	80	2025	0.954	0.951	0.964	0.955	1.577	8.586	4.416
0.7	Beta	160	4050	0.951	0.950	0.960	0.951	0.757	5.713	2.282
0.7	Beta	900	22780	0.947	0.948	0.952	0.948	-0.008	2.166	0.406
0.9	Beta	40	1200	0.952	0.951	0.977	0.952	2.552	14.474	17.584
0.9	Beta	80	2400	0.954	0.953	0.971	0.959	1.300	10.038	9.231
0.9	Beta	160	4799	0.947	0.947	0.961	0.952	0.674	7.078	4.728
0.9	Beta	900	26994	0.948	0.948	0.956	0.950	0.072	3.043	0.956

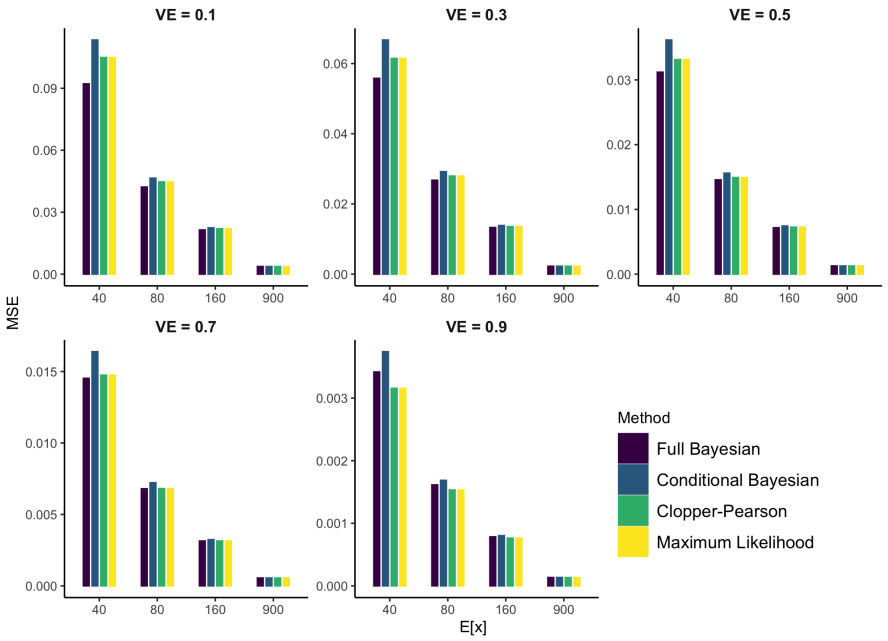
FB: Full Likelihood Bayesian, CB: Conditional Bayesian, CP: Clopper-Pearson frequentist, ML: Maximum Likelihood frequentist.

Analogamente alla stima intervallare, i miglioramenti più rilevanti si osservano quando il numero di casi è ridotto, con un guadagno decrescente al crescere del numero di infezioni, fino a raggiungere valori di MSE comparabili tra tutti i metodi.

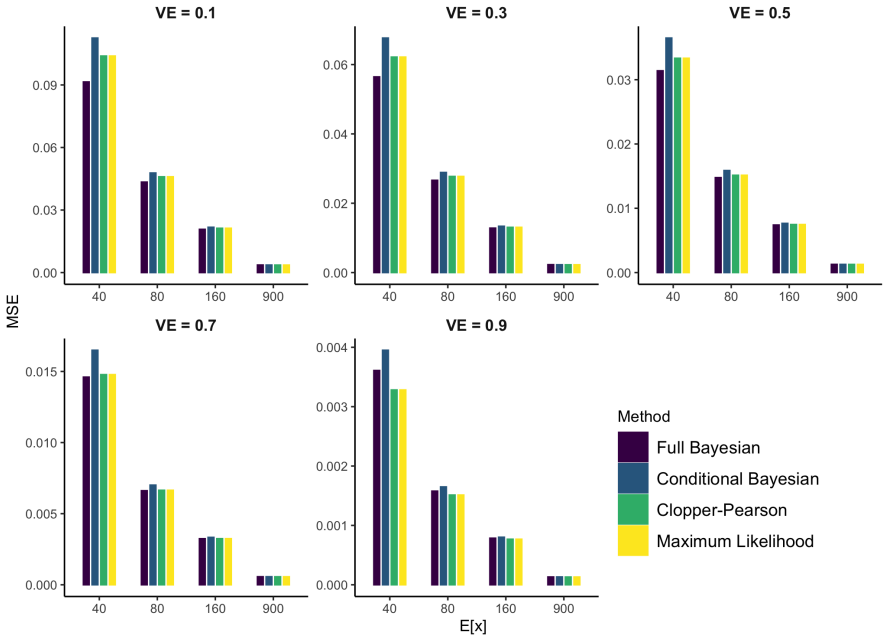
In ognuno degli scenari analizzati, il metodo FB porta sistematicamente a un MSE inferiore se confrontato con il metodo CB, tranne nelle scenario $\mathbb{E}[X_v + X_c] = 900$ in cui le medie a posteriori dei due approcci sono molto simili, indipendentemente dal valore di VE.

Escluso quest'ultimo scenario, a parità di numero totale di casi, la riduzione percentuale media dell'MSE decresce all'aumentare di VE rispetto a tutti i metodi concorrenti. Tuttavia, nel confronto con i metodi frequentisti, dalle tabelle 4.3 e 4.4 si evince che

l'errore quadratico medio del metodo FB tende ad aumentare per VE = 0.9, soprattutto se il numero di casi atteso è basso (piccoli campioni). Questa difformità tra il metodo bayesiano FB e i metodi frequentisti non deve però sorprendere. Infatti, poichè il metodo FB è sistematicamente migliore del metodo CB in termini di stima puntuale, anche il metodo bayesiano condizionato presenta la medesima discrepanza. Qualsiasi sia l'approccio bayesiano adottato, la distribuzione a posteriori e, di conseguenza, la media a posteriori combinano l'informazione derivante dai dati con quella derivante dalla prior, a differenza dell'approccio frequentista le cui inferenze dipendono esclusivamente dai dati. In grandi campioni, il contributo della distribuzione a priori diventa trascurabile e media a posteriori e MLE tendono ad equivalersi in termini di MSE [51]. In campioni di dimensione ridotta, invece, la media a posteriori presenta generalmente una varianza più bassa rispetto all'MLE, grazie al cosiddetto *effetto di shrinkage* della prior [4, 16], la quale può però introdurre bias. Nel caso specifico, se la prior è ben calibrata (il valore vero di VE è vicino a 0.3), il bias della media a posteriori si riduce e, congiuntamente a una varianza ridotta, porta a una diminuzione anche dell'MSE; se la prior è mal calibrata (il valore vero di VE è vicino a 0.9), il bias introdotto risulta significativamente maggiore del guadagno in varianza, peggiorando l'MSE della media a posteriori rispetto all'MLE. Infine, analogamente alle stime intervallari, il processo di reclutamento sembra non influenzare la riduzione percentuale dell'MSE dell'approccio FB rispetto agli altri metodi.



(a) Reclutamento uniforme



(b) Reclutamento beta

Figura 4.2: Errore quadratico medio (MSE) realtivo alla stima puntuale di VE. Le medie a posteriori sono usate come stime puntuali bayesiane, mentre l'MLE è usato per gli approcci frequentisti.

Tabella 4.3: Confronto tra l'approccio bayesiano con verosimiglianza completa e gli altri approcci in termini di stima puntuale: reclutamento uniforme.

VE	**R**	$\mathbf{E}[\mathbf{X_v}+\mathbf{X_c}]$	**N**	**MSE % reduction**			**Bias^2 % reduction**			**Variance % reduction**		
				CB	**CP**	**ML**	**CB**	**CP**	**ML**	**CB**	**CP**	**ML**
0.1	Uniform	40	697	18.753	12.162	12.162	81.452	43.188	43.188	14.893	11.514	11.514
0.1	Uniform	80	1393	9.257	5.583	5.583	82.968	41.738	41.738	7.016	5.255	5.255
0.1	Uniform	160	2786	4.522	2.461	2.461	80.323	39.109	39.109	3.223	2.256	2.256
0.1	Uniform	900	15668	-0.674	-1.043	-1.043	92.508	70.236	70.236	-0.898	-1.087	-1.087
0.3	Uniform	40	777	16.379	9.233	9.233	76.490	5.322	5.322	12.802	9.293	9.293
0.3	Uniform	80	1553	8.361	4.234	4.234	72.735	-1.045	-1.045	6.111	4.285	4.285
0.3	Uniform	160	3105	4.167	1.959	1.959	71.783	-3.398	-3.398	2.938	1.986	1.986
0.3	Uniform	900	17465	-0.157	-0.555	-0.555	88.474	28.257	28.257	-0.351	-0.565	-0.565
0.5	Uniform	40	879	13.759	5.851	5.851	67.732	-92.051	-92.051	10.446	6.900	6.900
0.5	Uniform	80	1757	6.646	2.462	2.462	68.177	-121.677	-121.677	4.806	3.005	3.005
0.5	Uniform	160	3514	3.584	1.231	1.231	63.337	-90.941	-90.941	2.477	1.562	1.562
0.5	Uniform	900	17465	-0.157	-0.555	-0.555	88.474	28.257	28.257	-0.351	-0.565	-0.565
0.7	Uniform	40	1014	11.379	1.510	1.510	51.254	-359.625	-359.625	8.303	4.585	4.585
0.7	Uniform	80	2028	5.850	0.164	0.164	46.775	-302.170	-302.170	3.953	2.056	2.056
0.7	Uniform	160	4055	2.898	0.007	0.007	48.378	-398.974	-398.974	1.910	0.913	0.913
0.7	Uniform	900	22808	0.256	-0.095	-0.095	66.691	-14211.721	-14211.721	0.126	-0.030	-0.030
0.9	Uniform	40	1202	8.652	-8.296	-8.296	27.193	-8816.166	-8816.166	6.144	1.874	1.874
0.9	Uniform	80	2403	4.316	-5.302	-5.302	23.156	-2962.486	-2962.486	2.784	0.852	0.852
0.9	Uniform	160	4806	2.302	-2.686	-2.686	21.837	-2505.196	-2505.196	1.476	0.519	0.519
0.9	Uniform	900	27030	0.292	-0.675	-0.675	21.668	-1880.439	-1880.439	0.118	-0.069	-0.069

FB:Full Likelihood Bayesian approach, CB: Conditional Likelihood Bayesian approach, CP: Clopper-Pearson frequentist approach, ML: Maximum Likelihood frequentist approach

Tabella 4.4: Confronto tra l'approccio bayesiano con verosimiglianza completa e gli altri approcci in termini di stima puntuale: reclutamento beta riscalato.

VE	**R**	$\mathbf{E}[\mathbf{X_v}+\mathbf{X_c}]$	**N**	**MSE % reduction**			**Bias^2 % reduction**			**Variance % reduction**		
				CB	**CP**	**ML**	**CB**	**CP**	**ML**	**CB**	**CP**	**ML**
0.1	Beta	40	696	18.717	11.966	11.966	80.007	40.845	40.845	14.682	11.298	11.298
0.1	Beta	80	1391	9.134	5.530	5.530	82.349	40.692	40.692	6.945	5.212	5.212
0.1	Beta	160	2782	4.584	2.458	2.458	75.765	32.313	32.313	3.122	2.237	2.237
0.1	Beta	900	15647	-0.323	-0.731	-0.731	96.899	83.992	83.992	-0.514	-0.764	-0.764
0.3	Beta	40	776	16.560	9.188	9.188	74.070	5.386	5.386	12.744	9.260	9.260
0.3	Beta	80	1551	7.814	4.000	4.000	76.383	-3.422	-3.422	5.775	4.051	4.051
0.3	Beta	160	3101	4.010	1.724	1.724	70.867	-1.997	-1.997	2.701	1.745	1.745
0.3	Beta	900	17443	-0.932	-1.328	-1.328	93.653	54.278	54.278	-1.118	-1.344	-1.344
0.5	Beta	40	878	13.967	5.854	5.854	66.436	-82.391	-82.391	10.567	6.949	6.949
0.5	Beta	80	1755	6.945	2.431	2.431	64.557	-94.881	-94.881	4.944	3.058	3.058
0.5	Beta	160	3510	3.299	0.968	0.968	61.781	-84.193	-84.193	2.171	1.312	1.312
0.5	Beta	900	19740	-0.149	-0.557	-0.557	70.447	-45.362	-45.362	-0.393	-0.525	-0.525
0.7	Beta	40	1013	11.453	1.173	1.173	49.559	-300.518	-300.518	8.232	4.514	4.514
0.7	Beta	80	2025	5.632	0.471	0.471	50.200	-450.805	-450.805	3.897	2.089	2.089
0.7	Beta	160	4050	2.825	0.086	0.086	48.153	-395.485	-395.485	1.861	0.975	0.975
0.7	Beta	900	22780	0.037	-0.471	-0.471	54.471	-493.851	-493.851	-0.143	-0.346	-0.346
0.9	Beta	40	1200	8.640	-9.984	-9.984	25.025	-2340.644	-2340.644	6.000	2.036	2.036
0.9	Beta	80	2400	4.309	-4.353	-4.353	25.154	-12930.425	-12930.425	2.883	0.831	0.831
0.9	Beta	160	4799	2.036	-2.262	-2.262	23.587	-10894.039	-10894.039	1.300	0.349	0.349
0.9	Beta	900	26994	0.404	-0.628	-0.628	20.237	-1777.886	-1777.886	0.238	0.006	0.006

FB:Full Likelihood Bayesian approach, CB: Conditional Likelihood Bayesian approach, CP: Clopper-Pearson frequentist approach, ML: Maximum Likelihood frequentist approach

4.4 Rivisitazione dello studio di Pfizer/BioNTech

In questa sezione la nuova metodologia proposta viene applicata allo studio clinico condotto da Pfizer/BioNTech sul vaccino anti-COVID-19 [12, 50], svoltosi tra il 2020 e il 2021. In Tabella 4.5 sono riportati i dati relativi all'intera popolazione del trial, suddivisi in diversi periodi e in differenti sottogruppi. L'analisi è condotta utilizzando l'approccio bayesiano con verosimiglianza completa (FB), l'approccio bayesiano condizionato (CB), la procedura di Clopper–Pearson (CP) e il metodo della massima verosimiglianza asintotica (ML).

Le informazioni relative alle dimensioni campionarie, al numero di infezioni e ai tempi di sorveglianza per ciascuna coorte sono disponibili pubblicamente [12, 50]. I dati riguardanti la durata massima a rischio D (espressa in anni) sono stati ottenuti sottraendo 28 giorni dalla durata totale del trial, calcolata a partire dalla data di inizio del reclutamento (27 luglio 2020) fino alla data di cut-off dei dati (14 novembre 2020 in Polack [12] e 13 marzo 2021 in Thomas [50]). Tale correzione è dovuta al protocollo dello studio, che considera un partecipante a rischio di infezione a partire da 7 giorni dopo la seconda dose di vaccino (somministrata 21 giorni dopo la randomizzazione).

Sebbene tutti i metodi forniscano stime puntuali comparabili dell'efficacia vaccinale (VE) nei sottogruppi considerati, emergono differenze significative nell'ampiezza dei corrispondenti intervalli al 95%. In particolare, l'approccio FB produce costantemente intervalli più stretti, suggerendo una maggiore precisione rispetto agli altri metodi. Come discusso nella sezione 4.3, tale vantaggio risulta più evidente nei sottogruppi più piccoli, ossia quelli con un numero ridotto di infezioni osservate, come la coorte brasiliana e i soggetti di età superiore ai 65 anni. In questi sottogruppi, dove il numero di infezioni è di circa 10 e 20 rispettivamente, gli intervalli di credibilità al 95% ottenuti con il metodo FB risultano più stretti di circa il 12% e 5% rispetto a CB, del 33% e 21% rispetto a CP e del 37% e 32% rispetto agli intervalli di confidenza basati su ML.

Un miglioramento più modesto, ma comunque significativo, della precisione si osserva nei sottogruppi con un numero moderato di infezioni, come il sottogruppo degli ispanici e quello maschile analizzati in Polack [50]. In questi casi, il metodo FB riduce l'ampiezza degli intervalli circa dell'1.5% rispetto al metodo CB e circa del 12% rispetto agli approcci frequentisti.

Infine, tutti i metodi producono intervalli praticamente identici quando applicati all'intera popolazione dello studio riportata in Thomas [12] (marzo 2021), dove il numero

di infezioni supera 900. Questo risultato è coerente con le simulazioni riportate nella sezione 4.3 e riflette la convergenza asintotica attesa delle diverse procedure statistiche all'aumentare del numero di eventi osservati.

Tabella 4.5: Confronto delle stime di VE con approcci bayesiani e frequentisti utilizzando i dati pubblicati [12], [50] del trial di Pfizer/BioNTech sul vaccino anti-COVID-19.

		Data			VE × 100 estimation				% Width reduction		
Ref	**Subgroup**	**Control**	**Vaccine**	**Duration**	**FB**	**CB**	**CP**	**ML**	**CB**	**CP**	**ML**
Thomas [50]	Overall March 2021	$n_c = 20713$ $x_c = 850$ $s_c = 6003$	$n_v = 20712$ $x_v = 77$ $s_v = 6247$	$D = 0.55$	91.27 (89.07,93.14)	91.26 (89.04,93.13)	91.30 (89.00, 93.20)	91.30 (89.01, 93.11)	0.22	3.04	0.64
Polack [12]	Overall Nov. 2020	$n_c = 17511$ $x_c = 162$ $s_c = 2222$	$n_v = 17411$ $x_v = 8$ $s_v = 2214$	$D = 0.21$	94.87 (90.38,97.63)	94.83 (90.33,97.63)	95.04 (90.00, 97.90)	95.04 (89.92, 97.56)	0.59	8.19	5.13
Polack [12]	Male	$n_c = 8762$ $x_c = 81$ $s_c = 1108$	$n_v = 8875$ $x_v = 3$ $s_v = 1124$	$D = 0.21$	96.00 (89.82,98.90)	95.93 (89.67,98.89)	96.35 (88.94, 99.26)	96.35 (88.44, 98.85)	1.46	12.01	12.74
Polack [12]	Hispanic or Latinx	$n_c = 4746$ $x_c = 53$ $s_c = 600$	$n_v = 4764$ $x_v = 3$ $s_v = 605$	$D = 0.21$	93.88 (84.18, 98.33)	93.78 (83.92, 98.29)	94.39 (82.68, 98.88)	94.39 (82.04, 98.25)	1.58	12.68	12.74
Polack [12]	Over 65	$n_c = 3880$ $x_c = 19$ $s_c = 511$	$n_v = 3848$ $x_v = 1$ $s_v = 508$	$D = 0.21$	93.30 (73.17,99.24)	92.92 (71.71,99.22)	94.71 (66.70, 99.87)	94.71 (60.45, 99.29)	5.10	21.32	32.79
Polack [12]	Brazil	$n_c = 1121$ $x_c = 8$ $s_c = 117$	$n_v = 1129$ $x_v = 1$ $s_v = 119$	$D = 0.21$	85.85 (38.09,98.49)	84.30 (29.59,98.36)	87.71 (8.33, 99.72)	87.71 (1.74, 98.46)	12.18	33.92	37.56

FB: Full Likelihood Bayesian, CB: Conditional Bayesian, CP: Clopper-Pearson frequentist, ML: Maximum Likelihood frequentist.

4.5 Conclusioni e sviluppi futuri

In questo lavoro è stato proposto un nuovo approccio bayesiano per la stima dell'efficacia vaccinale (VE). La principale innovazione risiede nella modellizzazione congiunta del numero di infezioni e dei tempi di sorveglianza totali di ciascun gruppo, ottenuta tramite un'approssimazione asintotica basata sul teorema limite centrale bivariato.
Un ampio studio simulativo, che ha esplorato diversi scenari di VE, processi di reclutamento e dimensioni campionarie, ha mostrato che l'approccio proposto fornisce intervalli di credibilità più stretti rispetto ai metodi bayesiani e frequentisti standard, senza perdita in termini di copertura, oltre a una riduzione dell'errore quadratico medio delle stime puntuali. Questo vantaggio è particolarmente evidente quando il numero di infezioni è basso o moderato, rendendo il metodo utile soprattutto nelle analisi per sottogruppi o in studi clinici di piccole e medie dimensioni.
Sebbene nel modello proposto i primi due momenti dei tempi di sorveglianza dipendano dal processo di reclutamento, questi parametri incogniti possono essere facilmente stimati all'interno dell'impostazione bayesiana. Di conseguenza, l'approccio non richiede assunzioni specifiche sulla distribuzione dei tempi di reclutamento, risultando altamente flessibile. Per semplicità, l'analisi si è concentrata su due processi stilizzati di reclutamento: un processo uniforme e un processo con tempi di reclutamento distribuiti secondo una Beta riscalata. Tuttavia, ci si aspetta che le prestazioni cambino solo marginalmente anche qualora si considerassero processi di reclutamento più realistici, come quelli proposti da Senn [45] o Anisimov [1].
Mentre nell'approccio bayesiano condizionato è disponibile una distribuzione a posteriori analitica per VE (grazie alla coniugatezza beta-binomiale), non esiste un'espressione in forma chiusa per l'approccio bayesiano con verosimiglianza completa, che richiede quindi l'utilizzo di metodi MCMC. Ciononostante, data la limitata complessità del modello, la stima a posteriori rimane altamente efficiente e non presenta difficoltà pratiche.
I risultati ottenuti possono avere importanti implicazioni per la progettazione di futuri studi clinici vaccinali di dimensioni intermedie (come quelli discussi da Ewell [11]), in particolare per quanto riguarda la determinazione della numerosità campionaria e le strategie di reclutamento.

Appendice A

Codici R-JAGS

A.1 Simulazione dei dati in R

```r

simula_trial <- function(n, D, B, lambda_c, VE, dist_flag, shape1,
    shape2) {

  # ---- RANDOMIZZAZIONE ----
  lambda_v <- (1 - VE) * lambda_c
  group <- rep(c("v", "c"), each = n/2)
  rates <- ifelse(group == "v", lambda_v, lambda_c)
  dist_flag <- tolower(as.character(dist_flag))
  dist_flag <- match.arg(dist_flag, c("uniform", "beta"))
  if (dist_flag == "uniform") {
    rec_times <- runif(n, 0, B)
  } else {
    rec_times <- rbeta(n, shape1 = shape1, shape2 = shape2) * B
  }
  TTE <- rexp(N, rates)

  # ---- TEMPI DI SORVEGLIANZA INDIVIDUALI ----
  surv_times <- pmin(D-rec_times,TTE)
  in_study <- rec_times <= D          # chi entra prima della fine

  # ---- DATI AGGREGATI PER GRUPPO ----
  s_v <- sum(surv_times[group == "v" & in_study])
```

```
  s_c <- sum(surv_times[group == "c" & in_study])
  x_v <- sum(group == "v" & (rec_times + TTE) <= D)
  x_c <- sum(group == "c" & (rec_times + TTE) <= D)
  n_v <- sum(group == "v" & in_study)
  n_c <- sum(group == "c" & in_study)

  return(list(s_v = s_v, s_c = s_c, x_v = x_v, x_c = x_c,
              n_v = n_v, n_c = n_c))
}
```

Listing A.1: Function R per la simulazione dei dati

A.2 Implementazione in R-JAGS del metodo bayesiano completo e dei metodi standard

```
library(rjags)
library(coda)

run_jags_inference <- function(x_v, x_c, s_v, s_c, n_v, n_c, D,
   n_iter = 20000, n_burnin = 2000, n_chains = 3, thin = 20) {

  # ---- MODELLO JAGS (FB) ----
  model_string <- "
  model {

    # ---- GRUPPO VACCINO ----

    # ---- Prior su E[min(T_v,C)] ----
    Em_v ~ dunif(0,D)

    # ---- Prior su P(T_v < C) ----
    p_v <- (1-VE)*p_c*Em_v/Em_c

    # ---- Prior su E[min(T_v,C)^2] ----
    varm_v ~ dunif(0,D^2)
    Em2_v <- Em_v^2 + varm_v

```

```
    #---- GRUPPO CONTROLLO ----

    # ---- Prior su E[min(T_c,C)] ----
    Em_c ~ dunif(0,D)

    # ---- Prior su P(T_c < C) ----
    p_c ~ dunif(0,1) #dbeta(0.1*Em_c/(1-0.1*Em_c),1)

    # ---- Prior su E[min(T_c,C)^2] ----
    varm_c ~ dunif(0,D^2)
    Em2_c <- Em_c^2 + varm_c

    # ---- VEROSIMIGLIANZA COMPLETA ----
    #(X_v, S_v) ~ normale bivariata
    mu_xv <- n_v*p_v
    mu_sv <- n_v*Em_v
    var_xv <- n_v*p_v*(1-p_v)
    var_sv <- n_v*(Em2_v - Em_v^2)
    cov_v <- n_v*p_v*(0.5*Em2_v/Em_v - Em_v)
    mu_sv_cond <- mu_sv + cov_v/(var_xv) * (x_v - mu_xv)
    var_sv_cond <- var_sv - cov_v^2/var_xv
    x_v ~ dbin(p_v,n_v) #dnorm(mu_xv, 1/var_xv)
    s_v ~ dnorm(mu_sv_cond, 1/var_sv_cond)

    # (X_c, S_c) ~ normale bivariata
    mu_xc <- n_c*p_c
    mu_sc <- n_c*Em_c
    var_xc <- n_c*p_c*(1-p_c)
    var_sc <- n_c*(Em2_c - Em_c^2)
    cov_c <- n_c*p_c*(0.5*Em2_c/Em_c - Em_c)
    mu_sc_cond <- mu_sc + cov_c/(var_xc) * (x_c - mu_xc)
    var_sc_cond <- var_sc - cov_c^2/var_xc
    x_c ~ dbin(p_c,n_c) #dnorm(mu_xc, 1/var_xc)
    s_c ~ dnorm(mu_sc_cond, 1/var_sc_cond)

    # ---- Prior Pfizer su VE ----
    theta_hat <- (1-0.3)/(2-0.3)
    a <- theta_hat/(1-theta_hat)
```

```
60      theta ~ dbeta(a,1)
61      VE <- 1 - theta/(1-theta)
62
63    }
64    "
65
66    jags_data <- list(
67      x_c = x_c,
68      x_v = x_v,
69      s_v = s_v,
70      s_c = s_c,
71      D = D,
72      n_v = n_v,
73      n_c = n_c
74    )
75
76    model <- jags.model(textConnection(model_string), data =
          jags_data,
77                        n.chains = n_chains, n.adapt = 5000)
78
79    update(model, n.iter = n_burnin) # scarta i primi n_burnin
          campioni a posteriori
80
81    samples <- coda.samples(model, variable.names = c("VE"), n.iter
          = n_iter, thin = thin)
82    ve_samples <- as.matrix(samples)[, "VE"] # combina tutte le
          catene in un'unica matric
83    ci_full <- quantile(ve_samples, c(0.025, 0.975))
84
85    # ---- MODELLO BETA-BINOMIALE (CB) ----
86    theta_hat <- (1-0.3)/(2-0.3)
87    a <- theta_hat/(1-theta_hat)
88    theta_samples <- rbeta(3000, a + x_v, 1 + x_c)
89    ve_samples_cond <- ((1-theta_samples)-theta_samples*s_c/s_v)/(1-
          theta_samples)
90    ci_cond <- quantile(ve_samples_cond, c(0.025, 0.975))
91
92    # ---- METODO DELLA MASSIMA VEROSIMIGLIANZA (ML) ----
```

```
  z <- qnorm(0.975,0,1)
  xv <- x_v + 0.5
  xc <- x_c + 0.5
  sv <- s_v + 0.5
  sc <- s_c + 0.5
  ci_ml <- c(1 - exp(log(xv*sc/(xc*sv)) + z*sqrt(1/xv + 1/xc)),
             1 - exp(log(xv*sc/(xc*sv)) - z*sqrt(1/xv + 1/xc)))

  # ---- METODO DI CLOPPER-PEARSON (CP) ----
  L <- qbeta(0.025, x_v, 1 + x_c)
  H <- qbeta(0.975, 1 + x_v, x_c)
  ci_cp <- c(1 - H/(1 - H)*s_c/s_v, 1 - L/(1 - L)*s_c/s_v)

  return(list(
    prob_VE_full = prob_VE_full,
    VE_CI_full = ci_full,
    VE_mean_full = mean(ve_samples),
    prob_VE_cond = prob_VE_cond,
    VE_CI_cond = ci_cond,
    VE_mean_cond = mean(ve_samples_cond),
    VE_CI_ml = ci_ml,
    VE_CI_cp = ci_cp,
    MLE_v = x_v/s_v,
    MLE_c = x_c/s_c,
    conv = max(gelman.diag(samples)$psrf), # Convergenza tra
        catene
    autcorr = autocorr.diag(samples)[2,"VE"], # Autocorrelazione
    ess = floor(effectiveSize(samples)["VE"]) # Campioni
        indipendenti
  ))
}
```

Listing A.2: Function R per l'implementazione dei 4 metodi statistici: FB,CB,CP,ML

Bibliografia

[1] Vladimir V Anisimov and Valerii V Fedorov. Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. *Statistics in medicine*, 26:4958–75, 11 2007.

[2] RE Blakesley, S Mazumdar, MA Dew, PR Houck, G Tang, CF Reynolds, and MA Butters. Comparisons of methods for multiple hypothesis testing in neuropsychological research. *Neuropsychology*, 23(2):255–264, 2009.

[3] E. Brittain and J. J. Schlesselman. Optimal allocation for the comparison of proportions. *Biometrics*, 38(4):1003–1009, 1982.

[4] George Casella and Roger L. Berger. *Statistical Inference*. CRC Press, 2024.

[5] CDC. *Principles of Epidemiology in Public Health Practice. An Introduction to Applied Epidemiology and Biostatistics*. U.S. Department of Health and Human Services, 2012.

[6] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

[7] JD Ciolino, C Wang, M Yu, et al. Guidance on interim analysis methods in clinical trials. *Contemporary Clinical Trials*, 129:107142, 2023.

[8] David R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–220, 1972.

[9] Jay Devore. *Probability and Statistics for Engineering and the Sciences*. Cengage Learning, 2014.

[10] Byar D. P. et al. Randomized clinical trials. perspectives on some recent ideas. *The New England Journal of Medicine*, 295(2):74–80, 1976.

[11] M Ewell. Comparing methods for calculating confidence intervals for vaccine efficacy. *Statistics in medicine*, 15:2379–92, 1996.

[12] Stephen J Thomas et al. Fernando P Polack. Safety and efficacy of the bnt162b2 mrna covid-19 vaccine. *The New England journal of medicine*, 383:2603–2615, 12 2020.

[13] Mauro Gasparini. *Modelli probabilistici e statistici*. C.L.U.T., 2014.

[14] Mauro Gasparini. On some modeling issues in estimating vaccine efficacy. *Pharmaceutical Statistics*, 9 2024.

[15] Edmund A. Gehan. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *Journal of Chronic Diseases*, 13(4):346–353, 1961.

[16] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. CRC Press, 3rd edition, 2013.

[17] B. George, S. Seals, and I. Aban. Survival analysis and regression models. *Journal of Nuclear Cardiology*, 21(4):686–694, 2014.

[18] M. Greenwood and G. U. Yule. The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general. *Proceedings of the Royal Society of Medicine*, 8(Sect Epidemiol State Med):113–194, 1915.

[19] Andrew P. Grieve. Idle thoughts of a 'well-calibrated' bayesian in clinical drug development. *Pharmaceutical Statistics*, 15:96–108, 3 2016.

[20] G. Hamra et al. Integrating informative priors from experimental research in epidemiologic studies. *Epidemiology*, 24(2):312–320, 2013.

[21] Julian P. T. et al. Higgins. The cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343:d5928, 2011.

[22] L. A. Kalish and C. B. Begg. Treatment allocation methods in clinical trials: a review. *Statistics in Medicine*, 4(2):129–144, 1985.

[23] J. M. Lachin. Statistical properties of randomization in clinical trials. *Controlled Clinical Trials*, 9(4):289–311, 1988.

[24] J. M. Lachin, J. P. Matts, and L. J. Wei. Randomization in clinical trials: conclusions and recommendations. *Controlled Clinical Trials*, 9(4):365–374, 1988.

[25] Curt D. Furberg Lawrence M. Friedman and David L. DeMets. *Fundamentals of Clinical Trials*. Springer, 2010.

[26] Roderick J Little. Calibrated bayes. *The American Statistician*, 60:213–223, 8 2006.

[27] Roderick J Little. Calibrated bayes, for statistics in general, and missing data in particular. *Statistical Science*, 26, 5 2011.

[28] Roderick J. Little. Calibrated bayes, an inferential paradigm for official statistics in the era of big data. *Statistical Journal of the IAOS*, 31:555–563, 11 2015.

[29] Mengya Liu, Qing Li, Jianchang Lin, Yunzhi Lin, and Elaine Hoffman. Innovative trial designs and analyses for vaccine clinical development. *Contemporary Clinical Trials*, 100:106225, 1 2021.

[30] Ira M. Longini M. Elizabeth Halloran and Jr. Claudio J. Struchiner. *Design and Analysis of Vaccine Studies*. Springer, 2010.

[31] Gianluca Mastrantonio. Dispense del corso di modelli statistici. Versione 1.1.1, 2023.

[32] J. P. Matts and R. B. McHugh. Analysis of accrual randomized clinical trials with balanced groups in strata. *Journal of Chronic Diseases*, 31(12):725–740, 1978.

[33] JH McDonald. *Handbook of Biological Statistics*. Sparky House Publishing, 3rd edition, 2014.

[34] Jozef Nauta. *Statistics in Clinical and Observational Vaccine Studies*. Springer International Publishing, 2020.

[35] R. Peto, M. C. Pike, P. Armitage, N. E. Breslow, D. R. Cox, S. V. Howard, N. Mantel, K. McPherson, J. Peto, and P. G. Smith. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. i. introduction and design. *British Journal of Cancer*, 34(6):585–612, 1976.

[36] Richard Peto and Julian Peto. Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society: Series A (General)*, 135(2):185–198, 1972.

[37] Larry K. Pickering, H. Cody Meissner, Walter A. Orenstein, and Amanda C. Cohn. Principles of vaccine licensure, approval, and recommendations for use. *Mayo Clinic Proceedings*, 95:600–608, 3 2020.

[38] Martyn Plummer. *rjags: Bayesian Graphical Models using MCMC*, 2023. R package version 4-14.

[39] S. J. Pocock. Allocation of patients to treatment in clinical trials. *Biometrics*, 35(1):183–197, Mar 1979.

[40] K. Qian et al. Weighted log-rank test for clinical trials with delayed effects. *Mathematics*, 10(15):2573, 2022.

[41] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.

[42] Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[43] Sheldon M. Ross. *Probabilità e statistica per l'ingegneria e le scienze*. Maggioli Editore, 2023.

[44] Donald B. Rubin. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, 12, 12 1984.

[45] Stephen Senn. Some controversies in planning and analysing multi-centre trials. *Statistics in Medicine*, 17:1753–1765, 8 1998.

[46] Stephen Senn. *Statistical Issues in Drug Development*. Wiley, 6 2021.

[47] Stephen Senn. The design and analysis of vaccine trials for covid-19 for the purpose of estimating efficacy. *Pharmaceutical statistics*, 21:790–807, 7 2022.

[48] R. Simon. Optimal two-stage designs for phase ii clinical trials. *Controlled Clinical Trials*, 10(1):1–10, 1989.

[49] StataCorp. *Introduction to group sequential designs*, 2022. Stata Adaptive Design Manual.

[50] Nicholas Kitchin et al. Stephen J Thomas, Edson D Moreira. Safety and efficacy of the bnt162b2 mrna covid-19 vaccine through 6 months. *The New England journal of medicine*, 385:1761–1773, 11 2021.

[51] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

[52] Qinyu Wei, Peng Wang, and Ping Yin. Confidence interval estimation for vaccine efficacy against covid-19. *Frontiers in Public Health*, 10, 8 2022.

[53] M. Zelen. The randomization and stratification of patients to clinical trials. *Journal of Chronic Diseases*, 27(7-8):365–375, 1974.